



NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
SCHOOL OF HEALTH SCIENCES
SCHOOL OF MEDICINE
DEPARTMENT OF PATHOPHYSIOLOGY
DIRECTOR PROFESSOR M. VOULGARELIS

**Development of an integrated data mining
toolkit for the investigation of fibromics data**

Dionysios Fanidis, MSc

Computational Biologist, Bioinformatician

DOCTORAL DISSERTATION

Athens 2024



**NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
SCHOOL OF HEALTH SCIENCES
SCHOOL OF MEDICINE
DEPARTMENT OF PATHOPHYSIOLOGY
DIRECTOR PROFESSOR M. VOULGARELIS**

**Development of an integrated data mining
toolkit for the investigation of fibromics data**

Dionysios Fanidis, MSc

Computational Biologist, Bioinformatician

DOCTORAL DISSERTATION

Athens 2024

Application Date: 9/12/2019

Three Member Committee Designation Date: 26/02/2020

Three-member Committee:

- 1. Athanasios Tzioufas, PhD**
Professor of Pathophysiology – Rheumatology, University of Athens
- 2. Dimitris Fotiadis, PhD**
Professor of Biomedical Technology, University of Ioannina
- 3. Vassilis Aidinis, PhD**
Researcher A', BSRC Alexander Fleming

Topic Designation Date: 21/05/2020

Dean of the School of Medicine: Professor Nikolaos Arkadopoulos

PhD Dissertation Deposition Date: 14/04/2024

Seven Member Committee:

- 1. Athanasios Tzioufas, PhD**
Professor of Pathophysiology – Rheumatology, University of Athens
- 2. Dimitris Fotiadis, PhD**
Professor of Biomedical Technology, University of Ioannina
- 3. Vassilis Aidinis, PhD**
Researcher A', BSRC Alexander Fleming
- 4. Evangelos Andreakos, PhD**
Researcher A', Director of the Center for Clinical, Experimental Surgery & Translational Research, Biomedical Research Foundation Academy of Athens
- 5. Antonios Sakellarios, PhD**
Assistant Professor, University of Patras
- 6. Andreas Goules, PhD**
Assistant Professor of Rheumatology, University of Athens
- 7. Ioanna Galani, PhD**
Researcher C', Biomedical Research Foundation Academy of Athens

Grade: Honors (Άριστα)

Ὄρκος του Ἱπποκράτη

Ὅμνυμι Ἀπόλλωνα ἰητρὸν, καὶ Ἀσκληπιὸν, καὶ Ὑγίαν, καὶ Πανάκειαν, καὶ θεοὺς πάντας τε καὶ πάσας, ἴστορας ποιούμενος, ἐπιτελέα ποιήσῃς κατὰ δύναμιν καὶ κρίσιν ἐμὴν ὄρκον τόνδε καὶ ξυγγραφὴν τήνδε.

Ἠγήσασθαι μὲν τὸν διδάξαντά με τὴν τέχνην ταύτην ἴσα γενέτησιν ἐμοῖσι, καὶ βίου κοινώσασθαι, καὶ χρεῶν χρηρίζοντι μετάδοσιν ποιήσασθαι, καὶ γένος τὸ ἐξ ωὔτεου ἀδελφοῖς ἴσον ἐπικρινέειν ἄρρεσι, καὶ διδάξειν τὴν τέχνην ταύτην, ἣν χρηρίζωσι μανθάνειν, ἄνευ μισθοῦ καὶ ξυγγραφῆς, παραγγελίης τε καὶ ἀκροήσιος καὶ τῆς λοιπῆς ἀπάσης μαθήσιος μετάδοσιν ποιήσασθαι υἱοῖσί τε ἐμοῖσι, καὶ τοῖσι τοῦ ἐμὲ διδάξαντος, καὶ μαθηταῖσι συγγεγραμμένοισί τε καὶ ὠρκισμένοις νόμῳ ἰητρικῷ, ἄλλῳ δὲ οὐδενί.

Διαιτήμασί τε χρήσομαι ἐπ' ὠφελείῃ καμνόντων κατὰ δύναμιν καὶ κρίσιν ἐμὴν, ἐπὶ δηλήσει δὲ καὶ ἀδικίῃ εἴρξῃ.

Οὐ δώσω δὲ οὐδὲ φάρμακον οὐδενὶ αἰτηθεὶς θανάσιμον, οὐδὲ ὑψηγήσομαι ξυμβουλίην τοιήνδε. Ὅμοίως δὲ οὐδὲ γυναικὶ πεσσοῦν φθόριον δώσω. Ἄγνωσ δὲ καὶ ὀσίως διατηρήσω βίον τὸν ἐμὸν καὶ τέχνην τὴν ἐμὴν.

Οὐ τεμέω δὲ οὐδὲ μὴν λιθιῶντας, ἐκχωρήσω δὲ ἐργάτησιν ἀνδράσι πρήξιος τῆσδε.

Ἐς οἰκίας δὲ ὀκόσας ἂν ἐσίω, ἐσελεύσομαι ἐπ' ὠφελείῃ καμνόντων, ἐκτὸς ἐὼν πάσης ἀδικίης ἐκουσίης καὶ φθορίης, τῆς τε ἄλλης καὶ ἀφροδισίων ἔργων ἐπὶ τε γυναικείων σωμάτων καὶ ἀνδρῶν, ἐλευθέρων τε καὶ δούλων.

Ἄ δ' ἂν ἐν θεραπείῃ ἢ ἴδω, ἢ ἀκούσω, ἢ καὶ ἄνευ θεραπείης κατὰ βίον ἀνθρώπων, ἃ μὴ χρή ποτε ἐκλαλέεσθαι ἔξω, σιγήσομαι, ἄρῆρα ἠγεύμενος εἶναι τὰ τοιαῦτα.

Ὅρκον μὲν οὖν μοι τόνδε ἐπιτελέα ποιέοντι, καὶ μὴ ξυγγέοντι, εἴη ἐπαύρασθαι καὶ βίου καὶ τέχνης δοξαζομένῳ παρὰ πᾶσιν ἀνθρώποις ἐς τὸν αἰεὶ χρόνον. παραβαίνοντι δὲ καὶ ἐπιποροῦντι, τάναντία τουτέων.

Curriculum Vitae

Full name: Dionysios Fanidis

Date of Birth: 08/10/1995

Nationality: Greek

CURRENT POSITION

10/2023 – today Special Scientist – Bioinformatics/Computational biology
Diagnostic Laboratory, biobank.cy Center of Excellence

STUDIES

- 2019 – today PhD in Bioinformatics/ Computational Biology
Development of an integrated data mining toolkit for the investigation of fibromics data
Department of Medicine, National and Kapodistrian University of Athens (UOA) & BSRC ‘Alexander Fleming’, Vari, Greece
- 2017 – 2019 International MSc in Molecular Biomedicine
Department of Medicine, UOA & BSRC ‘Alexander Fleming’
<http://www.molecularbiomedicine.gr/>
- 2013 – 2017 BSc in Biology
Faculty of Biology, UOA

WORK EXPERIENCE

- 2019 – today PhD in Bioinformatics / Computational Biology
Skills/experience obtained:
1. Application of machine learning algorithms
 2. Analysis of various omics data (microarrays, RNA-seq, single cell RNA-seq, 16S rRNA-seq)
 3. Database design, implementation and maintenance
 4. Development of complex web servers/applications
 5. Development of statistical pipelines for data integration
 6. Very good knowledge of R, python and Linux
- 2018 – 2019 Graduate Researcher in Bioinformatics (MSc diploma thesis)
Skills/experience obtained:
1. Design and execution of computational protocols and experiments
 2. Development of Bioconductor R packages
 3. Analysis of RNA-seq data
- 5/2018 – 8/2018 Graduate trainee in Bioinformatics (MSc 2nd rotation)
ChIP-seq data and transcription factor motif analysis.
- 2/2018 – 4/2018 Graduate trainee (MSc 1st rotation)
Investigation of the role of activin-A signaling in T cells in a mouse Multiple Sclerosis model. Biomedical Research Foundation Academy of Athens
- 9/2016 – 8/2017 Undergraduate Researcher (BSc thesis)
Study of the mechanisms of cellular proteostasis and antioxidant response during carcinogenesis. Faculty of Biology, School of Sciences, UOA

SELECTED PUBLICATIONS

1. **Fanidis D.**, P. Moulos and V. Aidinis (2021). *Fibromine is a multi-omics database and mining tool for target discovery in pulmonary fibrosis*. Scientific Reports 11(1): 21712. [PMID: 34741074](#)
2. **Fanidis, D.**, V. C. Pezoulas, D. I. Fotiadis and V. Aidinis (2023). *An explainable machine learning-driven proposal of pulmonary fibrosis biomarkers*. Computational and Structural Biotechnology Journal 21: 2305-2315. [PMID: 37007651](#)
3. Nikitopoulou, I.*, **D. Fanidis***, K. Ntatsoulis*, P. Moulos, G. Mpekoulis, M. Evangelidou, A. G. Vassiliou, V. Dimakopoulou, E. Jahaj, S. Tshipilis, S. E. Orfanos, I. Dimopoulou, E. Angelakis, K. Akinosoglou, N. Vassilaki, A. Tzouvelekis, A. Kotanidou and V. Aidinis (2021). *Increased Autotaxin Levels in Severe COVID-19, Correlating with IL-6 Levels, Endothelial Dysfunction Biomarkers, and Impaired Functions of Dendritic Cells*. International Journal of Molecular Sciences 22(18). [PMID: 34576169](#) [* equal contribution]
4. Barbayianni, I., P. Kanellopoulou, **D. Fanidis**, D. Nastos, E.-D. Ntouskou, A. Galaris, V. Harokopos, P. Hatzis, E. Tsitoura, R. Homer, N. Kaminski, K. M. Antoniou, B. Crestani, A. Tzouvelekis and V. Aidinis (2023). *SRC and TKS5 mediated podosome formation in fibroblasts promotes extracellular matrix invasion and pulmonary fibrosis*. Nature Communications 14(1): 5882. [PMID: 37735172](#)
5. Galaris, A., **D. Fanidis**, E. Tsitoura, P. Kanellopoulou, I. Barbayianni, K. Ntatsoulis, K. Touloumi, S. Gramenoudi, T. Karampitsakos, A. Tzouvelekis, K. Antoniou and V. Aidinis (2023). *Increased lipocalin-2 expression in pulmonary inflammation and fibrosis*. Frontiers in Medicine 10. [PMID: 37746070](#)
6. Galaris A.*, **D. Fanidis***, E.-A. Stylianaki, V. Harokopos, A.-S. Kalantzi, P. Moulos, A. S. Dimas, P. Hatzis and V. Aidinis (2022). *Obesity Reshapes the Microbial Population Structure along the Gut-Liver-Lung Axis in Mice*. Biomedicines 10(2). [PMID: 35203702](#) [* equal contribution]
7. Magkrioti, C., G. Antonopoulou, **D. Fanidis**, V. Pliaka, T. Sakellaropoulos, L. G. Alexopoulos, C. Ullmer and V. Aidinis (2022) *Lysophosphatidic Acid Is a Proinflammatory Stimulus of Renal Tubular Epithelial Cells*. International Journal of Molecular Sciences. [PMID: 35806457](#)
8. Zannikou, M., I. Barbayianni, **D. Fanidis**, T. Grigorakaki, E. Vlachopoulou, D. Konstantopoulos, M. Fousteri, I. Nikitopoulou, A. Kotanidou, E. Kaffe and V. Aidinis (2021). *MAP3K8 Regulates Cox-2–Mediated Prostaglandin E² Production in the Lung and Suppresses Pulmonary Inflammation and Fibrosis*. The Journal of Immunology 206(3): 607-620. [PMID: 33443087](#)
9. Panagopoulou M., **D. Fanidis**, V. Aidinis and E. Chatzaki (2021). *ENPP2 methylation in health and cancer*. International Journal of Molecular Sciences 22(21). [PMID: 34769391](#)
10. Panagopoulou, M., A. Drosouni, **D. Fanidis**, M. Karaglani, I. Balgkouranidou, N. Xenidis, V. Aidinis and E. Chatzaki (2022). *ENPP2 Promoter Methylation Correlates with Decreased Gene Expression in Breast Cancer: Implementation as a Liquid Biopsy Biomarker*. International Journal of Molecular Sciences 23(7). [PMID: 35409077](#)

LANGUAGES

- Greek: Native speaker
- English: Certificate of Proficiency in English; the University of Michigan
IELTS Academic (Grade 7.5)
- French: Delf B2: utilisateur indépendant - independent user

CONFERENCES

1. **Dionysios Fanidis**, Panagiotis Moulos, Vassilis Aidinis (2021). *Fibromine: a tool for centralized exploration and integration of pulmonary fibrosis omics data*. 71st National Conference of HSBMB, Athens, Greece
2. **Dionysios Fanidis**, Panagiotis Moulos (2019). *Combined Statistics for Differential Expression Analysis of RNA-Sequencing Data*. IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), Athens, Greece, pp. 170-173, <https://doi.org/10.1109/BIBE.2019.00038>

SUMMER SCHOOL ATTENDANCE

- 2019 Volos Summer School of Human Genetics
Genomic data handling and analysis.

Στην Έλενα, την σύντροφο της ζωής μου.
Στους γονείς και τα αδέρφια μου.

Θερμές ευχαριστίες στην επιτροπή μου, στους συναδέλφους
και εξαιρέτως στον Δρ. Αϊδίνη για την
αμέριστη υπομονή και στήριξη.

Πᾶσα δόσις ἀγαθὴ καὶ πᾶν δῶρημα τέλειον ἄνωθεν ἐστὶ,
καταβαῖνον ἐκ σοῦ τοῦ Πατρὸς τῶν Φώτων.

Table of Contents

Disclaimer	13
Abstract	14
Περίληψη	15
Abbreviations	17
1. Introduction	18
1.1 Idiopathic Pulmonary Fibrosis	18
1.2 Murine model of IPF	19
1.3 IPF-implicated molecules	19
1.3.1 Autotaxin, LPA and IPF.....	19
1.3.2 Lipocalin 2 and IPF.....	20
1.3.3 TKS5, podosomes and IPF.....	21
1.3.4 <i>MAP3K8</i> in inflammation and IPF.....	21
1.3.5 <i>COL1A1</i> and pulmonary macrophages	22
1.4 The microbiome in pulmonary pathologies	22
1.5 COVID-19 and IPF	23
1.6 Omics and IPF	23
1.6.1 Omics technologies in IPF	23
1.6.2 Setting the standards for transcriptomics data handling.....	24
1.7 Transcriptomics & IPF	25
1.7.1 Microarrays in IPF	25
1.7.2 IPF through next-generation sequencing	29
1.7.3 Pulmonary fibrosis at the single cell level	31
1.8 Integration efforts and meta-analysis	40
1.8.1 Challenges in data meta-analysis	40
1.8.2 Meta-analytic/ integration efforts in IPF.....	41
1.8.3 Machine learning in IPF.....	42
2. Materials & Methods	45
2.1 Transcriptomics data selection for Fibromine	45
2.2 Fibromine-hosted microarray data analysis	45
2.2.1 Data preprocessing.....	45
2.2.2 Quality control, curation and probe filtering.....	46
2.2.3 Gene level summarization and differential expression analysis	46
2.3 Other microarray data analyses	47
2.3.1 Kidney disease datasets.....	47
2.4 Fibromine-hosted RNA-seq data analysis	47

2.4.1 Reads alignment.....	47
2.4.2 Reads quantification, normalization and filtering.....	48
2.4.3 Statistical analysis.....	48
2.5 Other RNA-seq data analyses.....	49
2.5.1 Inflammatory diseases data.....	49
2.5.2 Quant-Seq data.....	49
2.6 TCGA data analysis.....	49
2.7 Microbiomics data analysis.....	50
2.8 scRNA-seq data analyses.....	50
2.8.1 GSE122960 data.....	50
2.8.2 COVID-19 data.....	51
2.8.3 Kidney disease data.....	52
2.8.4 Fibromine-hosted data.....	52
2.8.5 <i>LCN2</i> expression assessment.....	52
2.9 Weighted Gene Co-expression Network Analysis.....	53
2.10 Pathway analysis.....	53
2.11 Multiplex ELISA data analysis.....	54
2.12 Machine learning.....	54
2.12.1 Data pre-processing.....	54
2.12.2 Sex-specific expression examination.....	54
2.12.3 Tuning, training and evaluation.....	55
2.12.4 Model explanation.....	55
2.12.5 Biomarker lists comparison.....	55
2.13 Ranking aggregation.....	56
2.14 Text mining.....	56
2.15 Transcription factor analysis.....	56
2.16 Connectivity Map analysis.....	57
2.17 Fibromine-hosted proteomics data handling.....	57
2.18 Fibromine database creation.....	57
2.19 Protein-protein interaction networks.....	58
2.20 Accompanying wet laboratory experimentations.....	59
3. Results.....	59
3.1 A centralized resource for pulmonary fibrosis omics.....	59
3.1.1 Datasets collection.....	59
3.1.2 Transcriptomics datasets re-analysis.....	59
3.1.3 Data organization.....	60

3.1.4 Transcriptomics benchmarking.....	62
3.1.5 Interactive database access/use	63
3.2 Machine learning-based disease marker prioritization.....	67
3.2.1 Successful XGBoost separation of sample phenotypes	67
3.2.2 Identification and functional characterization of ML-utilized features.....	70
3.2.3 Disease targets prioritization.....	71
3.2.4 Prioritized features list comparison with previous biomarker lists	73
3.3 ENPP2 expression and methylation in cancer	75
3.3.1 ENPP2 methylation in cancerous and healthy tissue	76
3.3.2 ENPP2 expression and methylation in prostate cancer.....	79
3.3.3 ENPP2 expression and methylation in lung cancer	80
3.3.4 ENPP2 expression and methylation in hepatocellular carcinoma	82
3.3.5 ENPP2 expression and methylation in UALCAN database data.....	82
3.4 ENPP2 methylation as a liquid biomarker of breast cancer	83
3.4.1 ENPP2 methylation and expression in breast cancer.....	83
3.4.2 ENPP2 methylation and expression between breast cancer stages and types.....	85
3.4.3 ENPP2 methylation and expression between breast cancer types	87
3.4.4 ENPP2 methylation as measured in ccfDNA	87
3.5 Gut, liver and lung microbiome.....	89
3.6 Lipocalin-2 in fibrosis and steady state.....	92
3.6.1 Human <i>LCN2</i> expression patterns.....	92
3.6.2 Mouse <i>Lcn2</i> expression patterns	94
3.6.3 Genetic approach of <i>Lcn2</i> role in murine lung pathology.....	97
3.7 TKS5 and fibroblast migration.....	99
3.7.1 <i>TKS5</i> expression is upregulated during PF	99
3.7.2 Lung fibroblasts intrinsically create TGF β -induced TKS5 podosome rosettes	101
3.7.3 <i>Tks5</i> haploinsufficiency diminish BLM-induced lung fibrosis.....	104
3.7.4 <i>Tks5</i> haploinsufficiency diminish BLM-induced lung fibrosis.....	104
3.7.5 Inhibiting Src does reduce podosomes creation and diminish pulmonary fibrosis.....	107
3.8 MAP3K8 against fibrosis	112
3.8.1 Decreased <i>Map3k8</i> expression during pulmonary fibrosis	112
3.8.2 <i>Map3k8</i> exerts anti-inflammatory functions especially in macrophages	114
3.8.3 <i>Map3k8</i> -produced PGE2 attenuates pulmonary fibrosis.....	119
3.9 Dendritic cells and COVID-19	122
3.9.1 Increase in ENPP2 levels during COVID-19.....	122
3.9.2 Cellular pattern of ENPP2 expression in COVID-19	126

3.10 LPA and renal fibrosis	128
3.10.1 HKC-8 proteome changes induced by 176 stimulants.....	129
3.10.2 LPA-driven changes in HKC-8 secretome and proteome.....	129
3.10.3 HKC-8 responses to the other stimuli.....	131
3.10.4 Focus on cluster II stimulants.....	134
3.10.5 <i>In vitro</i> validation and dissection of LPA effects.....	134
4. Discussion	137
4.1 New IPF targets from public omics resources	137
4.2 Methylation-dependent transcriptional control of <i>ENPP2</i> during cancer	141
4.3 The lung microbiome	143
4.4 Fibrosis-related features	145
4.4.1 LCN2 and fibrosis.....	145
4.4.2 TKS5 and fibroblast invasiveness.....	146
4.4.3 MAP3K8 and fibrosis.....	148
4.5 Fibrosis and COVID-19	149
4.6 LPA in the kidneys	151
References	153
Appendix I – Supplementary figures	194
Appendix II – Supplementary tables	214
Original publications	237

Disclaimer

No wet laboratory experimentations were conducted by D. Fanidis as part of his PhD research.

All wet laboratory experimentations presented herein were performed by other laboratory members, as referenced in the respective publications.

Wet laboratory processes/results are presented in this manuscript and its accompanying texts, e.g. appendices, for purposes of experimental workflow completeness.

For a technical description of the wet laboratory methods included in this thesis, please, refer to respective publications.

Abstract

It is currently estimated that human species can be affected by more than 10,000 distinct rare diseases. Although each one concerns a small percentage of the world's population, together they can create a huge economic burden. Among this long list of diseases lies idiopathic pulmonary fibrosis, the most lethal of the idiopathic interstitial pneumonias. Until now there is no curative option other than the scarcely available lung transplantation opportunities. Thus, patients are treated with anti-fibrotic agents trading time for often severe side effects. Over the next decades, disease burden is estimated to increase due to a pattern of increasing prevalence and a bidirectional relationship to COVID-19, necessitating research intensification.

In this thesis, *in silico* methods have been utilized in order to propose novel disease biomarkers and curative options for pulmonary fibrosis. Towards that goal we have created Fibromine, a database of manually curated and consistently re-analyzed omics data, spanning two species and numerous experimental setups. Its contents are freely accessible via the homonym online application which can also be used for real-time data combination, statistical and visual exploration. Subsequently, Fibromine-proposed deregulated features between pulmonary fibrosis and control conditions were used for machine learning prioritization of disease biomarkers. A short and a long target lists were obtained from SHAP-based explanation of the models, containing both well established and interesting novel targets. Both gene lists were capable of separating pathological from steady state samples with an at least equal performance to that of previously proposed feature sets. Overall, computational analysis has proven capable of identifying new promising disease targets once applied on high quality data. Centralization of the latter is necessary for future studies acceleration, hypothesis formation and/or validation.

Moreover, various omics data have been processed for the investigation of pulmonary fibrosis-related molecules and conditions. Data analysis from control, hepatocellular adenocarcinoma, prostate, lung and breast cancer samples described a regulatory link between DNA methylation and *ENPP2* expression. Deregulated during cancer and measurable through non-invasive procedures, this regulatory connection was found of some prognostic value urging for more studies on the subject. Subsequently, microbiome exploration from murine gut, liver and lung suggested the existence of an axis connecting the three organs. Diet-induced obesity was proposed to cause dysbiosis, thus potentially impairing homeostatic balance. Next, investigation of the up-regulated lipocalin-2 expression during pulmonary fibrosis suggested the molecule in question as a biomarker for lung inflammation and respiratory functional status, motivating for further studies. In fibroblasts was found to be crucial for pulmonary fibrosis and extracellular matrix invasion. Importantly, targeting *Tks5* successfully attenuated pulmonary fibrosis via down-regulating podosomes formation, thus arising as a new promising therapeutic alternative. In another report, *MAP3K8* was revealed to have an anti-fibrotic character via mediating inflammation-related processes and Cox-2-mediated PGE₂ production. Furthermore, examination of COVID-19 data suggested the pathologic nature of ATX/LPA axis and its importance for dendritic cells homeostasis which is disrupted during SARS-CoV-2 infection. Last, a detailed study recorded the responses of HKC-8 cells post distinct treatments with 176 stimulants. Among them, LPA acted in a pro-inflammatory fashion indicating a possible pathogenic role in chronic kidney disease.

In total, the above projects created a central high quality database of pulmonary fibrosis omics data, proposed new target genes for the treatment of pulmonary fibrosis and have also delved deeper into obscure areas of the aforementioned and similar pathologies.

Περίληψη

Σύμφωνα με τις τελευταίες εκτιμήσεις το ανθρώπινο είδος μπορεί να επηρεαστεί από περισσότερες από 10.000 διακριτές σπάνιες ασθένειες. Παρότι κάθε μια αφορά ένα μικρό ποσοστό του παγκόσμιου πληθυσμού, στο σύνολό τους δημιουργούν ένα μεγάλο οικονομικό βάρος. Ανάμεσα τους βρίσκεται και η ιδιοπαθής πνευμονική ίνωση, η πιο θανάσιμη από τις ιδιοπαθείς διάμεσες πνευμονοπάθειες. Έως τώρα δεν υπάρχει άλλη θεραπευτική επιλογή πέραν των σπάνιων ευκαιριών μεταμόσχευσης πνευμόνων. Έτσι οι ασθενείς λαμβάνουν αντι-ινωτικούς παράγοντες ανταλλάσσοντας χρόνο με πολλές φορές σοβαρές παρενέργειες. Κατά τις επόμενες δεκαετίες το φορτίο της ασθένειας εκτιμάται πως θα αυξηθεί εξαιτίας ενός μοτίβου αυξανόμενου επιπολασμού και αμφίδρομης συσχέτισης με την νοσο COVID-19, επιβάλλοντας την εντατικοποίηση σχετικών ερευνών.

Σε αυτή τη διατριβή χρησιμοποιήθηκαν *in silico* μέθοδοι για την πρόταση καινοτόμων βιοδεικτών και θεραπευτικών επιλογών για την πνευμονική ίνωση. Προς αυτό τον στόχο δημιουργήσαμε το Fibromine, μια βάση δεδομένων από χειρωνακτικώς σχολιασμένα και συνεπώς επαν-αναλυμένα ωματικά δεδομένα που εκτείνονται σε δύο είδη και πολυάριθμες πειραματικές διευθετήσεις. Τα περιεχόμενα της είναι δωρεάν προσβάσιμα μέσω της ομώνυμης διαδικτυακής εφαρμογής η οποία μπορεί επίσης να χρησιμοποιηθεί για τον συνδυασμό, την στατιστική και οπτική εξερεύνηση των δεδομένων σε πραγματικό χρόνο. Ακολουθώντας, γενετικά στοιχεία που προτάθηκαν από το Fibromine ως απορρυθμισμένα μεταξύ πνευμονικής ίνωσης και συνθηκών ελέγχου χρησιμοποιήθηκαν για την ιεράρχηση βιοδεικτών χρήσει μηχανιστικής μάθησης. Μια βραχεία και μια μακρά λίστα στόχων αποκτήθηκε από εξήγηση των μοντέλων με τη χρήση τιμών SHAP, περιλαμβάνοντας τόσο καλώς ορισμένους όσο και ενδιαφέροντες νέους στόχους. Και οι δύο λίστες γονιδίων ήταν ικανές να διαχωρίσουν παθολογικά από υγιή δείγματα με τουλάχιστον ισάξια επίδοση με αυτή προηγούμενως προταθέντων συνόλων γενωμικών χαρακτηριστικών. Εν πολλοίς, η πληροφοριακή ανάλυση αποδείχθηκε ικανή στον εντοπισμό νέων πολλά υποσχόμενων παθολογικών στόχων, εφόσον εφαρμοστεί σε υψηλής ποιότητας δεδομένα. Η συγκέντρωση των τελευταίων είναι απαραίτητη για την επιτάχυνση μελλοντικών μελετών, καθώς και για τον σχηματισμό ή/και την επαλήθευση υποθέσεων.

Επιπροσθέτως, διάφορα ωματικά δεδομένα έτυχαν επεξεργασίας για την μελέτη συγγενών μορίων/συνθηκών της πνευμονικής ίνωσης. Ανάλυση δειγμάτων αναφοράς, προστατικού, πνευμονικού, μαστικού και ηπατικού καρκίνου περιέγραψε έναν ρυθμιστικό σύνδεσμο μεταξύ DNA μεθυλίωσης και της έκφρασης του *ENPP2*. Απορρυθμισμένη κατά τον καρκίνο και μετρήσιμη μέσω μη επεμβατικών μεθόδων αυτή η ρυθμιστική σύνδεση βρέθηκε να έχει κάποια προγνωστική αξία ωθώντας προς νέες έρευνες. Ακολουθώντας, εξερεύνηση του μικροβιώματος εντέρου, ήπατος και πνευμόνων ποντικών πρότεινε την ύπαρξη ενός άξονα μεταξύ των τριών οργάνων. Παχυσαρκία επαγόμενη από λιπαρή δίαιτα προτάθηκε πως προκαλεί δυσβίωση και πιθανώς διαταραχή της ομοιόστασης. Έπειτα, έρευνα της αυξημένης έκφρασης της lipocalin-2 κατά την πνευμονική ίνωση πρότεινε το εν λόγω μόριο ως έναν βιοδείκτη της πνευμονικής φλεγμονής και των πνευμονικών λειτουργιών, κινητοποιώντας προς περαιτέρω έρευνες. Επίσης, ο TSK5 διαμεσολαβούμενος σχηματισμός ποδοσωμάτων από πνευμονικούς ινοβλάστες βρέθηκε καίριας σημασίας για την πνευμονική ίνωση και την εισβολή στην εξωκυττάρια μήτρα. Η στόχευση του *Tks5* σταμάτησε επιτυχώς την πνευμονική ίνωση μέσω μειωμένου σχηματισμού ποδοσωμάτων, και έτσι αναδείχθηκε έως μια νέα πολλά υποσχόμενη θεραπευτική επιλογή. Σε άλλη αναφορά, το *MAP3K8* φανερώθηκε να έχει αντι-ινωτικό χαρακτήρα μέσω ρύθμισης φλεγμονωδών διαδικασιών και της Cox-2 διαμεσολαβούμενης παραγωγής PGE2. Επιπλέον, εξέταση COVID-19 δεδομένων πρότεινε την παθολογική φύση του άξονα ATX/LPA και τη σημαντικότητα του για την ομοιόσταση των δενδριτικών κυττάρων, η οποία διακόπτεται κατά τη διάρκεια της λοίμωξης με SARS-CoV-2. Τέλος, μια αναλυτική μελέτη κατέγραψε τις αποκρίσεις των HKC-8 κυττάρων σε 176 διακριτές επωάσεις.

Μεταξύ αυτών το LPA είχε προ-φλεγμονώδη δράση υποδεικνύοντας τον πιθανώς παθογονικό ρόλο του στην χρόνια νεφρική νόσο.

Εντέλει, οι προαναφερθείσες εργασίες δημιούργησαν μια κεντρική, υψηλής ποιότητας βάση ωματικών δεδομένων πνευμονικής ίνωσης, πρότειναν νέα γονίδια στόχους για την θεραπεία της, ενώ επίσης διείσδυσαν βαθύτερα σε αχαρτογράφητες περιοχές της ανωτέρω παθολογίας.

Abbreviations

Acellular ECM	aECM
Acute exacerbation	AE
Acute lung injury	ALI
Amplicon sequence variants	ASVs
Arachidonic acid	AA
Autotaxin	ATX
Bleomycin	BLM
Canonical correlation analysis	CCA
Connectivity map analysis	CMap
Consensus differentially abundant protein	cDEP
Consensus differentially expressed gene	cDEG
Consensus fold change	FC _{Consensus}
Control diet	CD
Differential expression analysis	DEA
Differentially expressed gene	DEG
Findability, Accessibility, Interoperability and Reusability	FAIR
Fold change	FC
Gene copy number	GCN
Gene expression omnibus	GEO
Gene ontology	GO
Gene significance	GS
Hematopoietic cells	HC
Hepatocellular carcinoma	HCC
High-fat diet	HFD
Highly variable genes	HVG
Idiopathic pulmonary fibrosis	IPF
Interstitial lung disease	ILD
Lung adenocarcinoma	LC
Lysophosphatidic acid	LPA
Lysophosphatidylcholine	LPC
Machine learning	ML
Matthew's correlation coefficient	MCC
Module eigengene	ME
Module membership	MM
Monte-Carlo cross validation	MCCV
Mutli-dimensional scaling	MDS
Mutual nearest neighbors	MNN
Next generation sequencing	NGS
Non-hematopoietic cells	nHC
Normalized enrichment score	NES
Pathway analysis	PA
Plasmacytoid dendritic cells	pDCs
Precision cut lung slices	PCLS
Primary normal human lung fibroblasts	NHLFs
Principal component analysis	PCA
Prostaglandin E2	PGE2
Prostate adenocarcinoma	PC
Pulmonary fibrosis	PF
RNA sequencing	RNA-seq
Shapley additive explanations	SHAP
Shared nearest neighbor	SNN
Single cell RNA sequencing	scRNA-seq
Topological overlap measure	TOM
Transcription factor	TF
Usual interstitial pneumonia	UIP

1. Introduction

1.1 Idiopathic Pulmonary Fibrosis

Idiopathic pulmonary fibrosis (IPF) is a fatal, progressive, fibrotic disease of the lungs with unknown etiology. It belongs to the family of interstitial lung diseases (ILDs) and is characterized by a usual interstitial pneumonia (UIP) pattern not attributable to any environmental or other known factors leading to ILD development (Podolanczuk, Thomson et al. 2023).

From a physiological point-of-view, disease progression leads to gradual decrease in lung function capacity, worsening of the related symptoms and finally death due to either respiratory failure or existing comorbidities (Lederer and Martinez 2018). The median life expectancy of a newly diagnosed person, is rather historically defined to span a period 3 to 5 years post diagnosis (Nathan, Shlobin et al. 2011). Nevertheless, per patient disease evolution is very heterogeneous, multifactorial and thus greatly unpredictable (Podolanczuk, Thomson et al. 2023).

There are multiple IPF risk factors, both genetic and environmental. Mutations of the *MUC5B* promoter (Seibold, Wise et al. 2011, Hancock, Hennessy et al. 2018), variants of *TOLLIP* (Noth, Zhang et al. 2013) and genetic changes near *AKAP13* gene 57 (Allen, Porte et al. 2017) are only some of the genetic links discovered. Most recently, peripheral leukocyte short telomere length has been associated with poor outcomes for immunosuppressed IPF patients (Newton, Zhang et al. 2018), while several studies have established an association between telomeric DNA maintenance and both the sporadic and the familial forms of IPF (Fingerlin, Murphy et al. 2013, Stuart, Choi et al. 2015, Zhang, Povysil et al. 2022). As far as the environmental factors are concerned, age (>60 years), male gender, present or past smoking and occupational exposures are some of the well-established risk factors (Baumgartner, Samet et al. 2000, Kaul, Lee et al. 2021).

Currently, IPF treatment is based on two anti-fibrotic agents, nintedanib and pirfenidone (King, Bradford et al. 2014, Richeldi, du Bois et al. 2014). Those two drugs are able to decelerate disease progression and lung failure gradual establishment. Nevertheless, neither of the two agents are capable of significantly relieving patient symptoms and unfortunately, most recipients discontinue their use due to emerging side effects (Podolanczuk, Thomson et al. 2023). In addition, the effects of pirfenidone and nintedanib administration are not well known and scientists have only recently beginning to assess their impact on acute exacerbations (Isshiki, Sakamoto et al. 2021, Urushiyama, Jo et al. 2022). Unfortunately, a usually precarious lung transplantation is the only curative option available and yet, it is inaccessible for the vast majority of IPF suffering individuals (Antoniou, Tsitoura et al. 2021),

From an epidemiological point-of-view, IPF is considered to be an orphan disease (ORPHA:2032) which, according to the European Union Regulation on Orphan Medicinal Products (1999), is any disease that affects less than a single individual per 2000 EU inhabitants. Within the first decades of the 21st century, there has been reported a sometimes dramatic increase both in the incidence and prevalence of the disease, which cannot be clearly attributed to a certain factor (Podolanczuk, Thomson et al. 2023).

Taking all the above into consideration, the identification of new disease targets and/or biomarkers is extremely crucial.

1.2 Murine model of IPF

Multiple animals have been used as a means of studying pulmonary fibrosis (PF). Domesticated animals, such as dogs and equines, spontaneously develop PF offering a more natural pathological context, but at the same time rendering experiments particularly time consuming (Tashiro, Rubio et al. 2017). More importantly, these animal pathologies does not closely resemble IPF. Alternative animal models can be established by artificially evoking PF in murine species using chemical agents, such as bleomycin and asbestos and even radiation (Tashiro, Rubio et al. 2017).

The bleomycin (BLM) based mouse model is the most widely used one mainly due to its easy use, fast development and reproducibility, along with its ability of replicating several of the IPF characteristics (Tashiro, Rubio et al. 2017, Barbayianni, Ninou et al. 2018). BLM is a mixture of glycopeptides with bacterial origin, clinically used in the battle against cancer (Barbayianni, Ninou et al. 2018). While there is a plethora of methods for BLM administration, previous findings from our laboratory have shown that oropharyngeal aspiration provides a set of advantages including reduction in experimental mortality (Barbayianni, Ninou et al. 2018).

The model unfolds into three stages: an early inflammatory, a fibrotic and a resolution phase, not seen during IPF. Although timing may differ according to dosage and route of administration, peak of fibrosis is usually reached by day 14 post-BLM and resolution starts from day 21 (Mouratis and Aidinis 2011). Through the secretion of pro-inflammatory and pro-fibrotic agents, BLM activates fibroblasts and causes, among others, collagen deposition in the lungs, thus replicating a set of important IPF milestones (Barbayianni, Ninou et al. 2018). Lack of alveolar epithelium hyperplasia and spontaneous resolution in young mice not repeatedly treated with BLM constitute some of the most important drawbacks of the model (Mouratis and Aidinis 2011, Barbayianni, Ninou et al. 2018).

1.3 IPF-implicated molecules

1.3.1 Autotaxin, LPA and IPF

Autotaxin (ATX; coded by *ENPP2*) is a glycoprotein of the extracellular space that catalyzes the production of lysophosphatidic acid (LPA) via lysophosphatidylcholine (LPC) hydrolysis (Ninou, Magkrioti et al. 2018, Magkrioti, Galaris et al. 2019). Previous findings from our laboratory indicated that ATX expression is necessary for embryonic development, as it regulates the proper formation of vasculature and the neuronal system (Fotopoulou, Oikonomou et al. 2010). LPA, the effector of ATX, can be found in most biological fluids, primarily following ATX expression pattern (Magkrioti, Galaris et al. 2019). Being functionally pleiotropic, LPA mediates several inflammation-implicated processes, such as stromal remodeling, immune system regulation and vascular homeostasis (Magkrioti and Aidinis 2013, Magkrioti, Galaris et al. 2019).

ENPP2, the ATX coding gene, is made out of 27 exons and one less introns and is located in the 8q24 chromosomal position, the epicenter of frequent cancer-related genetic changes (Yang, Lee et al. 2002, Brisbin, Asmann et al. 2011). *ENPP2* transcripts are subject to alternative splicing with the recorded variants having various stability and expression patterns (Giganti, Rodriguez et al. 2008). Apart from the splicing-related regulation, *ENPP2* expression has been witnessed to be regulated by epigenetic mechanisms (Parris, Kovács et al. 2014). Nevertheless, evidence regarding *ENPP2* methylation during disease is scarce with *ENPP2*

promoter being hypo-methylated during biliary atresia (Udomsinprasert, Kitkumthorn et al. 2017) and hyper-methylated in breast carcinoma (Parris, Kovács et al. 2014).

In several contexts, *ENPP2* its products act in a pro-pathologic fashion. Indicatively, increased ATX and LPA levels have been detected in many pathologies including IPF, while ATX inhibition in animal models diminished disease progression (Oikonomou, Mouratis et al. 2012, Ninou, Kaffe et al. 2018, Ninou, Magkrioti et al. 2018, Magkrioti, Galaris et al. 2019). Moreover, several pieces of evidence suggest an active role of both ATX and LPA in chronic kidney disease (CKD) (Geng, Lan et al. 2012, Zhang, Chen et al. 2016, Zhang, Wang et al. 2017), a pathology with a dual fibrosis-inflammatory character. Specifically, LPA has been witnessed to increase the expression of pro-fibrotic factors from kidney tubular epithelial cells (TECs), thus affecting their secretory signature (Pradere, Klein et al. 2007, Geng, Lan et al. 2012, Sakai, Chun et al. 2017). Increased quantities of ATX have also been described in cancer (Peyruchaud, Saier et al. 2020) and interestingly, elevated levels of ATX and LPA in the blood have been correlated with cancer invasiveness (Yang, Lee et al. 2002, Benesch, Tang et al. 2014). In addition, ATX over-expression has been located in cancerous tissues (Memet, Tsalkidou et al. 2018).

1.3.2 Lipocalin 2 and IPF

Lipocalin 2 (encoded by *Lcn2/LCN2*) is a 25kDa protein belonging to the lipocalin superfamily (Triebel, Bläser et al. 1992, Kjeldsen, Johnsen et al. 1993, Flower, North et al. 2000). Although all superfamily members share a highly conserved core tertiary structure that defines their ligand-binding site, that of LCN2 is shallower and adequately large to harbor and form bonds with macromolecules (Goetz, Holmes et al. 2002). LCN2 can be found in a monomeric or a dimeric form and characterization of the its forms in urine, potentially enables the distinction between kidney pathologies and urinary tract infections (Cai, Rubin et al. 2010, Mårtensson, Xu et al. 2012, Nickolas, Forster et al. 2012).

LCN2 has two cell surface receptors and interaction with one of them, 24p3R, has been interestingly witnessed to promote cancer cells apoptosis via iron depletion (Devireddy, Gazin et al. 2005). As far as LCN2 ligands are concerned, the protein's ability to bind bacterial siderophores, small secreted iron chelators (Hider and Kong 2010), has been accidentally revealed several years after its initial discovery (Goetz, Holmes et al. 2002). As iron is an essential element for both humans and all aerobic microorganisms with very specific exceptions (Weinberg 1997, Posey and Gherardini 2000), there is an ongoing "competition" between a host and its microbes for iron uptake. The latter attempt to hijack iron using siderophores and the former responds with LCN2 secretion in order to neutralize them (Fischbach, Lin et al. 2006, Xiao, Yeoh et al. 2017).

In the context of human lung pathology, LCN2 has been detected in several respiratory diseases including chronic obstructive pulmonary disease (COPD) (Eagan, Damås et al. 2010, Treekitkarnmongkol, Hassane et al. 2020), ARDS (Kangelaris, Prakash et al. 2015) and viral infections, such as SARS-CoV-2 and influenza A (Huang, Li et al. 2022). More importantly, bronchial epithelial cells have been witnessed to produce LCN2 during IPF (Ikezoe, Handa et al. 2014) which has been linked to both iron homeostasis (Ali, Kim et al. 2020) and microbiome regulation (Molyneaux, Cox et al. 2014), processes in which LCN2 is implicated as described above. Nevertheless, despite this large volume of data suggesting LCN2 implication in PF,

further research is necessary to validate previous findings and dissect the exact mechanism of LCN2 action during IPF.

1.3.3 TKS5, podosomes and IPF

Lung fibroblasts are central players in PF. They secrete various ECM components including collagen, as well as ECM-remodeling enzymes, such as metalloproteinases (MMPs) [4]. The exact cell type composition and transcriptional profile of pathogenic fibroblast populations has not yet been fully elucidated (see *1.5.3.3 Single cells of the fibrotic mesenchyme*), but pro-fibrotic signals including TGF β do contribute to their divergent behavior (Penke and Peters-Golden 2019). During PF, it has been proposed that lung fibroblasts accumulate to fibrotic areas as a response to several chemoattractant molecules (Tomasek, Gabbiani et al. 2002, Penke and Peters-Golden 2019). This fibrotic hallmark, can be partially supported by the ability of lung fibroblasts to invade the ECM, a phenomenon that has been witnessed in both IPF patients and animal models of the disease (Li, Jiang et al. 2011, Lovgren, Kovacs et al. 2011, Karvonen, Lehtonen et al. 2012).

Cell invasion is achieved by ECM lysis realized via invadopodia in cancer and podosomes in other cell types. Podosomes are built around an actin-enriched core structure surrounded by scaffold proteins (Murphy and Courtneidge 2011, Paterson and Courtneidge 2018). TKS5 (encoded by *SH3PXD2A*) is an SRC substrate and scaffold protein of podosomes and invadopodia discovered several years ago (Lock, Abram et al. 1998, Abram, Seals et al. 2003, Seals, Azucena et al. 2005). Interestingly, *in vivo* Tks5 knockdown resulted in impaired tumorigenicity of mouse fibroblasts and thus affecting not only metastasis but also tumor growth (Blouw, Seals et al. 2008, Blouw, Patel et al. 2015). In addition, Tks5 has been proposed as a prognostic marker in multiple cancer types including lung adenocarcinomas (Li, Chen et al. 2013).

Based on the above findings and given the fact that PF is the primary risk factor for lung cancer (Karampitsakos, Tzilias et al. 2017), there is a necessity for a thorough exploration of TKS5 expression/function during IPF.

1.3.4 MAP3K8 in inflammation and IPF

MAP3K8 (or else *TPL2*) is a member of the MAPK kinase kinase family tasked with the activation of multiple downstream molecules, such as ERK and MEK. It can be found in two protein isoforms, a shorter and a longer one, each made out of three distinct structural areas. From an evolutionary point-of-view, its' amino acid sequence is highly conserved between humans and rodents (Vougioukalaki, Kanellis et al. 2011). Although initially identified as a proto-oncogene, rodent *Map3k8* has been found to exert both pro- and anti-tumor activity in a context-dependent manner. As far as its' human homologue is concerned, it seems that a somatic mutation of the gene does not crucially affect cancer progression. On the other hand, increased levels of the MAP3K8 protein levels can have a pro-tumorigenic potential (Gantke, Sriskantharajah et al. 2011).

In the lung, *Map3k8* is expressed in high levels and has been shown to protect against inflammation (Watford, Wang et al. 2009) and carcinogenesis (Gkirtzimanaki, Gkouskou et al. 2013). Moreover, MAP3K8-inhibition has been assessed as a protective countermeasure against ventilator-induced lung injury with, what seems to be, species-specific results (Kaniaris, Vaporidi et al. 2014, Manitsopoulos, Aidinis et al. 2017). As PF is closely related to

lung tumors (Karampitsakos, Tzilas et al. 2017), *MAP3K8* should be examined as a molecule potentially implicated in both pathologies.

1.3.5 *COL1A1* and pulmonary macrophages

Type I collagen is by far the most abundant protein of the human body. Composed of two $\alpha 1$ and a single $\alpha 2$ chains, is the primary component of the ECM of multiple tissues, such as the muscles and the tendons. In addition, to the provision of mechanical strength as in the aforementioned examples, it also architecturally supports parenchymal organs (Stefanovic 2013). The first chain of the protein is encoded by *COL1A1*, a gene located on the reverse strand of the 17th chromosome. According to Ensembl (GRCh38:CM000679.2) *COL1A1* has 13 transcripts from which only two are transcribed into a protein sequence. *COL1A1* orthologues can be found in 252 species, including 23 primates and 24 rodents and rodent-like animals.

COL1A1 has been implicated in multiple pathologies, rendering it a “usual suspect” in many different human disease contexts. Structural and quantitative *COL1A1* mutations are responsible for a great number of osteogenesis imperfecta cases, a rare skeletal pathology (Forlino and Marini 2016). In addition, *COL1A1* can function in a pro-tumorigenic function, although its exact mechanism of action has not been fully elucidated (Li, Sun et al. 2022). In the respiratory system, gene’s mRNA has been detected in bronchoalveolar lavage (BAL) cells and has also been measured in increased quantities during IPF (Tsitoura, Wells et al. 2016, Tsitoura, Vasarmidi et al. 2019). Recently, at the single cell level, accumulating evidence suggest that macrophages contribute to tissue regeneration via collagen production (Pilling, Fan et al. 2009, Simões, Cahill et al. 2020), thus necessitating the exploration of their function in PF.

1.4 The microbiome in pulmonary pathologies

Nowadays, the microbiome populating the animal tissues is recognized as an important factor for the regulation of steady state and disease by affecting host processes such as immunity and inflammation (Cho and Blaser 2012, Belkaid and Hand 2014). Microbiome is a fragile entity that can be affected by many environmental and chemical agents, such as diet and prescribed medications, in turn affecting the host himself (Cho and Blaser 2012). Such changes in the composition of the microbiome lead to dysbiosis, a condition that has been connected to several human diseases, including obesity (Fan and Pedersen 2021).

Strikingly, dysbiosis can have across-organ effects with changes in the gut microbiome been linked to non-alcoholic fatty liver disease (NAFLD) (Zhu, Baker et al. 2013, Jiang, Wu et al. 2015). Apart from this gut-liver axis manifested between organs in anatomical contact, similar axes connecting far apart locations, such as the gut-brain, the liver-lung and the gut-lung axis have been proposed (Arteel 2020, Zhang, Li et al. 2020, Morais, Schreiber et al. 2021). The latter signaling route has also been related to asthma and COPD (Zhang, Li et al. 2020). Moreover, the lungs have recently been found to host their own microbiome which is found deregulated during disease, while specific bacteria have been associated with the mortality of IPF patients (Mackintosh, Desai et al. 2019). Changes in diet were observed to reduce the risk of COPD in a potentially metabolic manner, thus suggesting an active connection between brain, lung and liver mediated by microbial metabolites (Young, Hopkins et al. 2015). Collectively, microbiome is a significant player in mammal physiology, but its exact role during pathology manifestation and progression has not been clearly illustrated.

1.5 COVID-19 and IPF

Firstly described in Wuhan, China as a pneumonia of “unknown cause”, corona virus disease 2019 (COVID-19) quickly evolved into a pandemic that dramatically affected the lives of millions of individuals all over the world (Hu, Guo et al. 2021). This new disease is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and has a median infection age around 50 years (Hu, Guo et al. 2021). Post-infection symptoms differ with age and pre-existing comorbidities and include, among others, fever, fatigue, dry cough, headache and chest pain, while anosmia, ageusia and other related disorders have been self-reported (Hu, Guo et al. 2021, Osuchowski, Winkler et al. 2021).

Hypoxaemia is one of the common and severe COVID-19 symptoms which can lead to an oxygenation impairment condition known as acute respiratory distress syndrome (ARDS). Roughly half of the people suffering by severe COVID-19 develop ARDS which in turn can lead to lung fibrosis (Osuchowski, Winkler et al. 2021). Fibrosis hallmarks such as epithelial-to-mesenchymal transition and increased TGF- β signaling have been observed in severe COVID-19 cases, justifying a high rate of PF among COVID-19 autopsies (Osuchowski, Winkler et al. 2021). In addition, scRNA-seq revealed many similarities in the transcriptional profile of IPF and COVID-19 (Bharat, Querrey et al. 2020, Wu, Yu et al. 2020). In parallel, patients suffering from ILDs are more vulnerable to a severe COVID-19 manifestation and poor outcomes following SARS-CoV-2 infection (Drake, Docherty et al. 2020, Esposito, Menon et al. 2020, Beltramo, Cottenet et al. 2021). Collectively, COVID-19 and IPF, although different in nature are highly entangled pathologic conditions necessitating their joint research.

1.6 Omics and IPF

Omics technologies is an umbrella term covering all high-throughput assays that can simultaneously measure molecules of the same type from a biological sample (Conesa and Beck 2019). Due to the fact that they are able to assess the quantity of all or nearly all elements of a given space, such as DNA in transcriptomics, proteins in proteomics or metabolites in metabolomics, omics technologies can provide a holistic examination of any given biological condition (Conesa and Beck 2019). Such means of molecular spaces exploration create huge datasets which in turn require complex computational methods for their handling, storage and analysis (Costa 2014).

Historically beginning from the examination of isolated types of molecules (single-omics), multiple molecular entities can nowadays be measured in parallel in the same sample (multi-omics) creating an ever spherical view of a system (Conesa and Beck 2019). Although, these multi-omic datasets require analysis in an integrative fashion, the scientific community is many times witnessing the uneven evolution of certain analytical tools in specified niches of research, thus unconsciously hindering inter-laboratory data sharing, comparison and interpretation (Manzoni, Kia et al. 2018). A possible reason for such cases may be the expensive and expertise-demanding nature of creating such multi-omics data collections, the abundant presence of which would eventually create the necessity for integrative analysis pipelines development.

1.6.1 Omics technologies in IPF

Various omics technologies have been recruited for IPF research. While transcriptomics is certainly the most widely used discipline spanning from microarrays and RNA-seq (Vukmirovic and Kaminski 2018) to scRNA-seq experiments (see 1.6.3 *Pulmonary fibrosis at*

the single cell level), other levels of molecular organization have also been investigated. Briefly, in the quest for PF biomarkers multiple studies have examined the proteome of various sampling sites including lung, plasma, serum and BAL fluid (Khan, Dasgupta et al. 2021). Genomics have also played a crucial role in identifying several disease-implicated genes such as *TERT* and *MUC5B* (Kan, Shumyatcher et al. 2017). In addition, epigenomics revealed DNA methylation changes in the lungs and fibroblasts of IPF patients potentially contributing to the disease progression (Kan, Shumyatcher et al. 2017).

In this thesis, focus is primarily given on transcriptomics datasets and secondly on proteomic ones as these two modalities occupy the vast majority of IPF omics space. The following sections attempt to shape the PF universe as seen through a transcriptomics and proteomics lens, as well as through several dataset integration attempts.

1.6.2 Setting the standards for transcriptomics data handling

Following data accumulation from the first high-throughput molecular assays, the scientific community realized the need for a new set of guidelines regarding data collection, handling, storage and report. Thus, to address limitations in microarray data handling, Brazma and colleagues (2001) proposed the Minimum Information About a Microarray Experiment (MIAME), a set of guidelines aiming to facilitate the replication, interpretation and automated meta-analysis/mining of microarray data (Brazma, Hingamp et al. 2001). Being platform independent, content- and not format- restricting, MIAME succeeded in the organization of microarray datasets in public repositories, such as the Gene Expression Omnibus (GEO) (Barrett, Wilhite et al. 2013), in an at least adequately annotated manner. In the following years, as next generation sequencing (NGS) gradually supplanted microarrays, new guidelines tailored to this new technology were once again required. To address this gap, Functional Genomics Data Society (FGED) published in 2012 the Minimal Information about a high-throughput Nucleotide SEQuencing Experiment (MINSEQE), aiming to the improvement of multi-model high-throughput datasets integration (<https://zenodo.org/record/5706412>).

A retrospective evaluation of guidelines implementation inferred that publicly accessible repositories, including GEO (Barrett, Wilhite et al. 2013), ArrayExpress (Rustici, Kolesnikov et al. 2013), DDBJ (Kodama, Mashima et al. 2012) and the later created Sequence Read Archive (SRA) (Kodama, Shumway et al. 2012), have embraced both MIAME and MINSEQE guidelines, yet with varying compliance rates (Rustici, Williams et al. 2021). Unfortunately, although the primary goal of meta-data deposition was largely achieved, there were indications that data sharing continues to lack consistency, thus rendering meta-analysis a rather tedious process. More specifically, analysis of publicly available datasets deposited from 2009 up to 2013 in ArrayExpress, GEO or SRA revealed that deposited data lacked crucial information concerning variables such as treatment conditions, biological replicates, library preparation and data processing protocols (Rustici, Williams et al. 2021). Results of the aforementioned review also suggest that in the same period, RNA-seq and microarray datasets had, on average, 2.8 and 3.6 biological replicates, respectively. As a result, many RNA-seq datasets can support very limited statistical analyses.

In 2014, a workshop held in Netherlands set the foundations for a set of principles that followed promise the increase of Findability, Accessibility, Interoperability and Reusability (FAIR) of scholarly data (Wilkinson, Dumontier et al. 2016). According to the observations of Wilkinson and colleagues (2016), the existing data ecosystem follows a decentralized

trajectory characterized by datasets deposition in smaller repositories, thus intensifying the problem of data discovery and reusability. For contexts like that of IPF, FAIR principles are not usually met as raw data, metadata and/or the algorithms/pipelines used for their analysis are not often carefully recorded. As expected, this holds true especially for datasets published prior to the definition of the aforementioned guidelines.

In total, while the aforementioned endeavors achieved in standardizing some of the high-throughput (meta-) data handling processes, both proper collection and data re/meta-analysis remain a great challenge.

1.7 Transcriptomics & IPF

Transcriptomics is the omics discipline targeting on the bulk analysis of RNA molecules. Multiple technologies have been developed for accomplishing such a feat, but microarrays and RNA sequencing are the most popular (Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials et al. 2012). Being already in use from the 1980s and developed to operate in a highly parallel fashion in the second half of the 1990s, microarrays offered from 2001 to 2006 the first genome wide parallel analysis of RNA molecules (Goodwin, McPherson et al. 2016, Levy and Myers 2016). Although nowadays surpassed by 2nd and 3rd generation sequencing technologies, microarrays had offered a cheaper and more robust solution to many DNA and RNA focusing applications for many years (Levy and Myers 2016). One of the major microarray limitations emerges from their probe-based way of function which preclude their use for data-driven discovery studies (Manzoni, Kia et al. 2018). On the other hand, RNA-sequencing, a 2nd generation (aka Next Generation Sequencing; NGS) sequencing technique, provides an unbiased means of RNA assessment in an ever reducing cost (Levy and Myers 2016, Shendure, Balasubramanian et al. 2017).

Without neglecting its advantages, RNA-sequencing is a bulk sequencing technique produces an average gene expression profile across vast amounts of cells, thus hindering the discovery of subtle transcriptome differences between different tissue locations and more importantly cell types (Olsen and Baryawno 2018). Designed to sequence a small number of rare cells, the first developed single cell RNA-sequencing pipeline (Tang, Barbacioru et al. 2009) paved the way for the rapid development of multiple sequencing protocols and *in silico* techniques with an ever increasing resolution (Svensson, Vento-Tormo et al. 2018).

In the following section, we will discuss publications that have used high-throughput techniques, attempting a thorough up-to-date recording of the progress made. Research deliverables exclusively utilizing low yielding techniques for IPF transcripts assessment, such as real time PCR, are reviewed elsewhere (Vukmirovic and Kaminski 2018).

1.7.1 Microarrays in IPF

Following the evolution of wet laboratory transcriptomics techniques, gene expression in IPF was initially assessed using microarrays. The first studies took place more than 20 years ago and revealed extensive differences in the transcriptome of IPF patients and BLM-treated rodents compared to their respective controls (Kaminski, Allard et al. 2000, Zuo, Kaminski et al. 2002). From that point on, IPF research community has witnessed a progressive accumulation of wet laboratory studies revolving around microarray-reported single deregulated genes including but not limited to *MMP7* (Cosgrove, Schwarz et al. 2002, Zuo, Kaminski et al. 2002, Gharib, Altemeier et al. 2013), *MMP3* (Yamashita, Dolgonos et al. 2011), *IGFBP-3/-5* (Pilewski, Liu et al. 2005), osteopontin (Pardo, Gibson et al. 2005), *CRLF1* (Kass,

Yu et al. 2012), *SI00A2* (Richards, Kaminski et al. 2012), *MMP19* (Yu, Kovkarova-Naumovski et al. 2012), *EGFR* (Tzouvelekis, Ntoliou et al. 2013), *LCLAT1* (Huang, Mathew et al. 2014), *VCAMI* (Agassandian, Tedrow et al. 2015), *FKBP10* (Staab-Weijnitz, Fernandez et al. 2015), *SIPL* (Huang, Berdyshev et al. 2015), caveolin-1 (Wang, Zhang et al. 2006), *PTPN11* (Tzouvelekis, Yu et al. 2016) and *DIO2* (Yu, Tzouvelekis et al. 2018) were continuously published.

Apart from single molecules, microarrays were also used to address changes in signaling pathways during fibrosis. To begin with, examination of the effects of IPF drivers, such as *TGF β* , led to the emergence of several important features including *PEDF*, *SMURF2*, *BMP4* and *NOX4* (Cosgrove, Brown et al. 2004, Renzoni, Abraham et al. 2004, Hecker, Vittal et al. 2009). In addition, macrophage-secreted syndecan-2 was distinguished among several *TGF β* signaling inhibitors as an anti-fibrotic in IPF lungs (Shi, Gochuico et al. 2013). In addition, *TNF α* effects were examined in liaison with those of *IFN γ* in order to assess the importance of Fas in the apoptosis-survival tradeoff of lung fibroblasts (Wynes, Edelman et al. 2011). At the same time *TNF α* has been implicated in regulating the expression of pathologic genomic features in human airway smooth muscle cells (Knobloch, Lin et al. 2013). Re-analysis of already published microarray data revealed Wnt pathway as one of the most significant for PF progression, with *LRP5* receptor levels pointing to the need of Wnt evaluation in conjunction with *TGF β* activation levels (Lam, Herazo-Maya et al. 2014). Validating those findings, the enhanced Wnt/ β -catenin pathway observed in *ABCG2+* mesenchymal progenitor cells (MPC) has been connected to various pro-fibrotic changes in the circulatory system (Gaskill, Carrier et al. 2017). Extending the findings of pre-clinical trials, *IL-4* and *IL-13* signaling pathways were found to induce similar transcriptional signatures in monocytes, macrophages and fibroblasts, while *IL-13Ra2* was discovered to act as a decoy receptor in the latter cell type (Chandriani, DePianto et al. 2014). Motivated by the synergistic action of *TNF α* signaling and heat shock on the airway epithelium, a research team extended their research to include clinically relevant wounds in the equation. As a result, they revealed increased expression of *FGF1* in response to the synergistic action of heat shock and epithelium wounding, while *FGF1* and *HSP70* were also found over-expressed in the fibrotic lung (Scheraga, Thompson et al. 2016). With a clear interest in the alveolar epithelium, Wasnick and colleagues (2019) identified and thoroughly examined *Notch1* signaling in both human and murine alveolar epithelial cells. Interestingly, enhancement of the *Notch1* signaling during fibrosis was held responsible for impaired epithelial cells turnover and the abnormal production of surfactant protein, thus providing an attractive target for pharmacological intervention (Wasnick, Korfei et al. 2019).

As ageing is closely related to the development of PF (Yanagihara, Sato et al. 2019) several studies have employed transcriptomics assays in order to investigate this interlink. First, following an observation of abnormal mitochondria in alveolar epithelial type II cells, decreased expression of *PINK1* has been shown to promote fibrosis via affecting mitophagy in the aging lung (Bueno, Lai et al. 2015). Moreover, as ageing-related diseases are characterized by an aberrant re-activation of developmental pathways in response to injuries, Chanda and colleagues (2016) have identified and transcriptionally characterized a population of mesenchymal stromal cells that seemed to participate in such a mechanism. Indeed, *TGF β 1* and *SHH* signaling pathways were found to suppress *FGF10* expression in the aforementioned

cell type of IPF subjects, while enrichment for developmental pathways further supported their working hypothesis (Chanda, Kurundkar et al. 2016).

Deciphering myofibroblast/IPF fibroblasts origin and modus operandi is under active research for many years. In 2007, microarrays were used to investigate the mechanisms underlying myofibroblasts pathogenicity (Kabuyama, Oshima et al. 2007), while almost in parallel, the transcriptome of the same cell type was assessed under different culturing conditions in an attempt to witness any matrix effect on transcriptional control (Larsson, Diebold et al. 2008). To unbiasedly characterize the full spectrum of RNAs produced by activated fibroblasts Emblom-Callahan and colleagues (2010) juxtaposed primary fibroblasts from IPF and control individuals, demonstrating changes in genes related to processes such as protein turnover, metabolic pathways and loss of cellular homeostasis (Emblom-Callahan, Chhina et al. 2010). Supplementing previous findings, suppression of interferon-stimulating genes was proposed as a key characteristic of pathogenic lung fibroblasts (Lindahl, Stock et al. 2013), while increased *FOXF1* expression has been proposed not as a pro-pathogenic but as a rather compensatory mechanism impaired in the fibrotic milieu (Melboucy-Belkhir, Pradère et al. 2014). In an attempt to illuminate extracellular matrix (ECM) – fibroblast interactions, cells were cultured in decellularized matrix isolated from either IPF or control individuals (Parker, Rossi et al. 2014). Results suggested that ECM is more important in shaping cultured fibroblasts transcriptome irrespective of their cellular origin and that reduced expression of miR-29 is responsible for increased ECM proteins concentration. Moreover, examination of cells upon isolation and after several culture passages, revealed that the *in vivo* expression profile of both IPF and control fibroblasts was gradually lost during cell culturing and that several of their *in vivo* transcriptomics differences were replaced by others of *in vitro* origin (Rodriguez, Emblom-Callahan et al. 2018). Last, microarrays have been applied for fibroblastic transcriptome evaluation following treatment with various agents such as TGF β , IL1, EGF and small molecule inhibitors (Bradley and Barrett 2017). Re-analysis of these data revealed that TGF β 1-induced *FZD7* is a central node to the regulation of multiple pro-fibrotic features, such as *ACTA2* and *COL1A1* (Guan and Zhou 2017). Apart from the above characteristics, fibrosis-promoting fibroblasts are highly migratory and invasive, a set of features proposed to be mediated by low levels of *USP13* (Geng, Huang et al. 2015) and high levels of methylation-controlled *BMPER* (Huan, Yang et al. 2015).

Microarrays were also applied for the characterization and distinction of IPF types/endotypes and patients affected by various comorbidities. In 2006, transcriptomics differences related to specific biological processes and cell types were revealed between slow and rapid PF progressors (Selman, Pardo et al. 2006) and a year later, transcriptomics differences between control, familial and sporadic interstitial idiopathic pneumonia were reported (Yang, Burch et al. 2007). As far as comorbidities are concerned, different gene signatures were indeed witnessed between PF cases with or without pulmonary arterial hypertension (Mura, Anraku et al. 2012). Differential expression of cilium genes efficiently separated two molecular IPF/UIP sub-phenotypes with an increased expression in the former associated with a more intense honeycombing pattern (Yang, Coldren et al. 2013). More importantly, in-parallel analysis of transcriptional and clinical data revealed that *RXFP1* expression levels could be used to predict patients' response to relaxin-based therapies (Tan, Tedrow et al. 2016).

In addition, microarrays provided an opportunity to examine the etiology of IPF acute exacerbations (AE). Although several hundred deregulated genes were identified between

stable and AE-hit patients, including *CCNA2* and α -defensins among the top up-regulated, neither an infectious nor an inflammatory cause was revealed (Konishi, Gibson et al. 2009). Even more interestingly, gene expression patterns during an AE were more similar to those of stable IPF samples compared to control lungs.

PF transcriptome was also set against that of other pathologies. For instance, microarray data suggested that progressive and fibrotic pulmonary sarcoidosis is more closely related to hypersensitivity pneumonitis than to IPF due to the first's transcriptome enrichment in immune system related genes (Lockstone, Sanderson et al. 2010). In order to pinpoint a biomarker signature detectable in peripheral blood, Rosas and colleagues (2008) examined the abundance of multiple plasma proteins and identified a combinatorial signature of five, able to distinguish patients from control individuals (Rosas, Richards et al. 2008). Moreover, MMP7 and MMP1 levels were not only sufficient for separating IPF from COPD and sarcoidosis cases, but their encoding genes were also found over-expressed in the lungs via microarrays further corroborating the importance of these two molecules. Transcriptome comparison between samples of pulmonary arterial hypertension and secondary pulmonary hypertension IPF patients pinpointed several differences including some of sex-specific nature (Rajkumar, Konishi et al. 2010). Microarray analysis of systemic sclerosis samples revealed unique molecular profiles both in lung tissue and fibroblasts when compared to IPF (Hsu, Shi et al. 2011). Non-specific interstitial pneumonia (NSIP) is another lung pathology whose transcriptome has been compared to that of IPF (Cecchini, Hosein et al. 2018). More specifically, NSIP and IPF have been validated as two molecularly distinct pathologies, while researchers have also pinpointed potential biomarkers capable of supporting the aforementioned distinction. In parallel, detection of a common signature between IPF and NSIP established a potential link of the former with the latter.

Although most PF-related microarrays focus on protein coding RNAs, there are also others examining non-coding transcripts, such as micro (miRNA) and circular (circRNA) RNA populations. The first such publication reported a set of 46 differentially expressed non coding RNAs, from which *let-7d* was distinguished as a molecule controlling the levels of *ACTA2*, *S100A4* and *SFTPC* (Pandit, Corcoran et al. 2010). A year later expression of mRNA and miRNA pairs was used to distinguish IPF from control individuals and IPF from NSIP patients (Cho, Gelinas et al. 2011). Interestingly, miR-23 cluster was proposed to control epithelial-to-mesenchymal transition, as well. Milosevic and colleagues (2012) examined the regulatory effects of TGF β on the expression and by extension the action of many miRNAs during PF, while they also proposed miR-154 being a regulator of fibroblast behavior in IPF (Milosevic, Pandit et al. 2012). miRNAs were once more proposed as mediators of TGF β signaling when another research group placed miR-145 as a regulatory node downstream of TGF β necessary for the overexpression of α SMA during PF (Yang, Cui et al. 2013). Apart from validating the deregulation of previously reported miRNAs, Huleihel and colleagues (2014) provided proof that *let-7d* induces the production of pro-fibrotic molecules, such as *ACTA2* and fibronectin, in primary fibroblasts further strengthening the hypothesis of miRNAs implication in fibrotic-related processes (Huleihel, Ben-Yehudah et al. 2014). On top of previous findings, miR-130b-3p was found under-expressed in IPF lung samples and subsequent wet laboratory experiments suggested that it is involved in not only fibroblast activation but also in epithelial-to-mesenchymal transition deregulation, thus promoting tissue fibrosis (Li, Geng et al. 2016). Last but not least, a specialized microarray assay identified 67 deregulated plasma circRNAs

participating in significant signaling pathways such as that of TGF β , VEGF and Wnt (Li, Wang et al. 2018).

Biomarker hunting is another research field that was boosted by transcriptomics technologies. Initially, transcriptomics evidence suggested that peripheral blood transcriptome is not only able to distinguish IPF patients from control individuals, but can also reflect disease burden (Yang, Luna et al. 2012). A year later, a 52 gene expression signature established through peripheral blood mononuclear cells (PBMCs) was used to predict IPF poor outcome (Herazo-Maya, Noth et al. 2013). Attempting a per time-point comparison of a rat BLM model with the actual IPF disease, Bauer and colleagues (2014) identified a set of biomarkers that can be used to classify samples into fibrotic or steady state. Fascinatingly, *CTHRC1*, a gene that emerged years later at the single cell level as characteristic of pathologic cell types (Tsukui, Sun et al. 2020) was listed among the markers (Bauer, Tedrow et al. 2014). Furthermore, gene expression patterns were used to identify two distinct types of IPF lung lesions, while *MMP3* and *CXCL13* protein product levels were successful in evaluating patients aggregate fibrotic burden (DePianto, Chandriani et al. 2015). In a screening of exclusively Japanese lung samples, *DSG3* was proposed as a new biomarker of idiopathic interstitial pneumonia (Horimasu, Ishikawa et al. 2017). That same year, microarray analysis of control fibroblasts and of cells from slow and rapid progressing IPF subjects led to the suggestion of *CCL8* as a biomarker of differential diagnosis and patient survival (Lee, Cheong et al. 2017). Interestingly, parallel examination of microarray and mass spectrometry data prompted for the use of groups instead of single biomarkers for lung disease differential diagnosis (Zhao, Yin et al. 2017). In addition to the aforementioned sampling sites, BAL fluid has been used for biomarker discovery, leading to the proposal of a 9-genes signature predictive of IPF mortality, especially when combined with clinical data. More importantly, discovery of this motif indicated that pulmonary basal cells may play an active role in IPF (Prasse, Binder et al. 2018).

Last, microbiome plays a significant role in many pathogenic contexts, such as in human metabolic diseases (Fan and Pedersen 2021), and IPF is no exception. In a study from Huang and colleagues (2017) examining PBMC transcriptome data in parallel with 16S rRNA, microbiome-host interaction was proposed to affect progression-free survival and fibroblast behavior in IPF patients via pattern recognition receptors and immune system signaling pathways (Huang, Ma et al. 2017). In addition, persistent dysbiosis of the lower airways has been suggested to act as a continuous source of alveolar injury potentially supporting the local fibrotic milieu during IPF (Molyneaux, Willis-Owen et al. 2017).

1.7.2 IPF through next-generation sequencing

RNA-sequencing does offer a number of advantages over microarrays, such as the disengagement from a pre-defined set of RNA-detecting probes and the support of a greater set of applications including the examination of alternative splicing. For example, in one of its first uses in the context of IPF, RNA-sequencing not only revealed a set of approximately 870 deregulated genes between fibrotic and control lungs, but it also identified another set of 440 features subjective to alternative splicing that included *COL6A3* and *POSTN* could be found (Nance, Smith et al. 2014). RNA splicing was again examined two years later in COPD, IPF and control lung samples where *NUMB* and *PDGFA* p53/hypoxia pathway-related genes were found more intensely spliced in pathological versus healthy samples (Kusko, Brothers et al. 2016). Expression analysis of the same data supported the existence of common mechanisms in both disease contexts. Moreover, novel differentially expressed alternative transcripts have

been identified between IPF and acute lung injury or control cases, as part of a general transcriptomics characterization of the three phenotypes (Sivakumar, Thompson et al. 2019).

Introduction of RNA-seq may have provided new viewing angles over transcriptome, but it did not ostracize its predecessor technology which was often used as a golden standard for comparison purposes. Thus, RNA-seq findings were cross-validated with those of previously published microarrays to corroborate the increased abundance of senescence biomarkers in the IPF lungs, with *CDKN2A* expression positively correlating with pathology severity (Schafer, White et al. 2017). Nance and colleagues (2014) have also used previously published array datasets to compare their RNA-seq-reported deregulated genes (Nance, Smith et al. 2014). From a more technical aspect Vukmirovic and colleagues (2017) showed that high-throughput sequencing of RNA molecules from archived Formalin-Fixed Paraffin-Embedded (FFPE) IPF tissues is not only plausible but also highly concordant with microarray application on fresh frozen samples (Vukmirovic, Herazo-Maya et al. 2017). For validation purposes, RNA-seq and microarrays were even applied on the same samples revealing the nowadays widely acceptable strong correlation of the two methods' results (Kusko, Brothers et al. 2016).

As with microarrays, RNA-sequencing was used multiple times for the examination of single molecule effects and the characterization of specific groups of cells. For example, NGS enabled the unbiased transcriptome evaluation of a new population of mesenchymal cells proposed to give birth to pathogenic lung fibroblasts (Xia, Bodempudi et al. 2014). Moreover, inspired by the re-analysis of an older microarray experiment which revealed *TAZ* as a fibroblast-implicated gene, RNA-seq examination of a *TAZ*-siRNA cell line suggested that connective tissue growth factor (CTGF) and collagen type I are regulated by *TAZ* (Noguchi, Saito et al. 2017). Motivated by the clusterin's dual role and its cellular localization-dependent expression pattern, Habel and colleagues (2017) re-examined data from a previously characterized mesenchymal population (Xia, Bodempudi et al. 2014) and identified a significant down-regulation of clusterin in fibrotic SSEA4⁺ cells (Habel, Camelo et al. 2017). Among the deregulated transcripts of the same cells, *S100A4* was one of the top over-expressed, while downstream experiments implied that *S100A4* is responsible for the fibrogenicity of mesenchymal progenitor cells (Xia, Gilbertsen et al. 2017). *BMP3* was also found deregulated during idiopathic interstitial pneumonias and wet laboratory experimentations provided hopeful data about its use as an IPF therapeutic target (Yu, Gu et al. 2017). Last, miRNAs have also been examined by RNA-seq, with Wang and colleagues very recently proposing that the peptide drug DR3penA does alleviate PF via hindering miR-23b-5p expression (Wang, Deng et al. 2023).

Examination of signaling pathways is yet another application that microarrays and NGS naturally have in common. Transcriptomics evaluation of lung epithelial cells treated or not with TGF- β 1 and E2 demonstrated for the first time that the latter regulates chromatin remodeling pathways in a TGF- β 1-dependent fashion (Smith, Moreno et al. 2018). Several other processes negatively regulated by the two aforementioned agents. Focusing in human pulmonary fibroblasts signaling, Mukherjee and colleagues (2019) revealed that *PTGER2* is the most highly expressed prostaglandin receptor in IPF patients. In addition, *PTGER2* expression was reduced in IPF compared to control fibroblasts suggesting an important role in the fibrotic context (Mukherjee, Sheng et al. 2019).

Adopting an alternative point-of-view, Luzina and colleagues (2018) examined whether normal appearing IPF tissue does differ from profoundly scarred areas or healthy lungs. Interestingly, based on RNA-seq data they supported that non-scarred IPF areas are actively implicated in disease progression, as they are enriched for various extracellular matrix-, immune- and inflammation-related mRNAs, similarly to scarred ones, when compared to healthy controls (Luzina, Salcedo et al. 2018). In parallel, actively fibrotic tissue areas differed from the non-scarred ones in terms of epithelial mucociliary mRNA molecules, thus indicating differences of the epithelium between the two tissue types.

Except for fibroblasts, macrophages have also interested scientists for their potential role in PF. Following certain indications about accumulation of activated macrophages in the lower parts of the respiratory tree, RNA-seq was used to characterize macrophages isolated from whole BAL fluid (Lee, Arisi et al. 2018). More specifically, and validated by downstream experimentations, the chemokine- and cytokine-signaling enriched list of differentially expressed genes (DEGs) suggested that alveolar macrophages are indeed involved in pro-inflammatory and tissue remodeling processes.

1.7.3 Pulmonary fibrosis at the single cell level

scRNA-seq studies have been primarily used for the examination of both fibrosis and steady state at a cellular level. Most have focused on a specific lung compartment, such as the epithelium or the stroma while others adopted a more holistic description of the lung.

1.7.3.1 Single cells of the lung epithelium

To begin with the lung epithelium, it is known that alveolar epithelial cells type 2 (AT2) trans-differentiate to type 1 cells (AT1) in order to replenish tissue lost due to lung injury. Impairment of this physiological process is nowadays thought to lead to the emergence of pathological cell populations which in turn contribute to the development and the establishment of various lung fibrotic cases. Adding to this working hypothesis, although examining a very small number of cells, one of the first scRNA-seq analyses of the IPF lung was able to identify disease specific epithelial cell populations and analyze them transcriptionally so as to identify ‘intermediate’ subpopulations characteristic of the fibrotic phenotype (Xu, Mizuno et al. 2017). During the same year, the first lung injury response mechanisms shared between epithelial cell types begun to emerge (Xi, Kim et al. 2017). In another set of single cells, sub-clustering of alveolar epithelial type 2 (AT2) cells revealed a specific group that expressed fibrosis-related genes and was located at fibrotic niches of the lung (Joshi, Watanabe et al. 2019). Enriching the ‘impaired regeneration’ hypothesis, Wnt/ β -catenin has been proposed to induce AT2 cells proliferation in response to lung injury (Riemondy, Jansing et al. 2019). Once completed, AT2 cell type expansion is halted by an increase in TGF β that places AT2 cells in cycle arrest. Subsequently, withdrawal of both the above signals leads to efficient trans-differentiation into AT1 cells. During PF, deregulation of this physiological process maintains high TGF β levels, thus placing all proliferating AT2 cells in a pro-fibrotic, terminal cell cycle arrest status (Riemondy, Jansing et al. 2019).

As it has been repeatedly reported, such as in (Xu, Mizuno et al. 2017, Morse, Tabib et al. 2019, Habermann, Gutierrez et al. 2020), the proportion of basal, ciliated and goblet cells and that of AT2 and AT1 cells has been found elevated and degraded in IPF individuals, respectively (Adams, Schupp et al. 2020). Focus on basal cells by the same study, described for the first time an IPF-specific pro-fibrotic and senescent population termed ‘aberrant

basaloid' cells, a finding which was further validated by re-analysis of an independent scRNA-seq dataset (Reyfman, Walter et al. 2018). In another synchronous study, two populations of basaloid lung cells were described: a CD66+ secretory primed (SPB) and a multi-potent one (Carraro, Mulay et al. 2020). During IPF, basal cells undergo a biased expansion towards the secretory type which is located in MUC5B-rich honeycomb regions and is primed for differentiation into secretory cell types.

Following their previous hypothesis of impaired alveolar epithelial cells differentiation (Riemyndy, Jansing et al. 2019) researchers from the same lab have provided evidence implicating a KRT8^{hi} KRT18^{hi} TGFβ^{hi} cluster of epithelial cells in the fibrotic process (Jiang, Gil de Rubio et al. 2020), further supporting the existence of a unique IPF AT2 cell state (Reyfman, Walter et al. 2018). More specifically, these cells were accumulated at fibroblastic foci and KRT8 expression peaked during the initial steps of epithelial differentiation, suggesting once again an impaired transition from AT2 to AT1 phenotype. Although not any across-species cellular comparison was presented, an independent paper supported the existence of a Krt8⁺ TGFβ-activated AT2 intermediate cell state, corroborating a common pathogenic mechanism in both model and actual disease (Strunz, Simon et al. 2020). Another piece of evidence was provided by Kobayashi and colleagues (2020) who by examining murine lung cells that significantly overlapped with those of (Riemyndy, Jansing et al. 2019) they have identified a Cldn4+Krt19+Sfn+ Pre-Alveolar Type-1 transitional cell state (PATS). PATS were further separated into two groups: a Ctgf+ and a Lgals3+ one, with the former presenting several similarities to the Krt8+ cells of (Strunz, Simon et al. 2020). Trajectory analysis applied on the above populations suggested that Ctgf+ PATS differentiate into AT1 cells traversing through the Lgals3+ group, with TP-53 being an essential signaling component of the process. Re-analysis of human data from (Habermann, Gutierrez et al. 2020) identified PATS-like cells during IPF thus strengthening the hypothesis of the accumulation of an intermediate state in the fibrotic milieu (Kobayashi, Tata et al. 2020).

In support of an impaired epithelial turnover, Habermann and colleagues (2020) proposed that a population of IPF KRT5-/KRT17+ epithelial cells emerge from a transitional AT2 cell state and acts in a pro-fibrotic manner via production of various ECM components, including collagen (Habermann, Gutierrez et al. 2020). A cell group transcriptionally intermediate of AT2 and AT1 cells that resembled the transient cluster described by (Xu, Mizuno et al. 2017) was also pinpointed (Habermann, Gutierrez et al. 2020). At the same time, re-analysis of (Reyfman, Walter et al. 2018, Morse, Tabib et al. 2019) datasets yielded closely related results.

Time-point analysis of the BLM model provided evidence supporting the accumulation of another AT2-AT1 intermediate cell population in the fibrotic lungs, termed Alveolar Differentiation Intermediate (ADI) (Strunz, Simon et al. 2020). These Krt8+ cells expressed p53 and NfκB in addition to senescence markers and they were also detected in other disease models and in IPF cases (Strunz, Simon et al. 2020). More importantly, trajectory analysis pinpointed activated AT2 and fully differentiated AT1 cells as the source and the terminal developmental stage of Krt8+ cells, respectively. In comparison with other similar cell types, ADI were transcriptionally related to KRT5-/KRT17+ and aberrant basaloid cells described by (Adams, Schupp et al. 2020, Mayr, Simon et al. 2021). Interestingly, ADI cells were also detected in control lung tissue where they function as homeostatic AT2 and AT1 intermediates.

From a more physiological perspective and based on observations from murine lungs, it was suggested that the increase of mechanical tension during PF accompanied by an impaired AT2 to AT1 trans-differentiation process is sufficient to trigger an epithelial Tgf β activation which in turn targets adjacent mesenchymal cells leading to severe fibrosis (Wu, Yu et al. 2020). Furthermore, inhibiting Cdc42 expression impaired AT2 to AT1 trans-differentiation thus resulting in a human-like periphery to center lung fibrosis establishment.

One of the factors that most probably contributes to a faulty lung epithelial regeneration is senescence. In pursuit of this hypothesis, bulk NGS-inspired signatures of senescent epithelial cells anthologized by *in vitro* experiments were used for the characterization of scRNA-seq data (DePianto, Vander Heiden et al. 2021). As a result, a novel senescent population of basal cells was discovered, closely resembling the transcriptional profile of terminally differentiated cells found in stratified epithelia. In comparison to previous publications, this new group of cells has homologues in (Reyfman, Walter et al. 2018, Adams, Schupp et al. 2020, Habermann, Gutierrez et al. 2020) datasets, but it is distinct from the aberrant basaloid cells of (Adams, Schupp et al. 2020, Habermann, Gutierrez et al. 2020) potentially due to differences in the applied computational approaches. Another piece of evidence regarding the importance of senescence in PF, came along with the information that early p53 activation targeting can block fibrogenesis in murine lungs (Yao, Guan et al. 2020). In this study, AT2 senescence was deemed a sufficient process in fibrosis establishment.

PF diagnosis and monitoring can be an extremely complex process and thus, discovery of effective non-invasive methods is highly anticipated. Working towards that goal, Mayr and colleagues (2021) have attempted to connect changes in plasma proteome with pathologic changes in lung physiology so as to propose a simple assay for assessing PF progression. According to their findings, CRTCA1 decreased peripheral levels positively correlated with multiple lung function measurements in ILD patients. Interestingly, CRTCA1 origin was traced back to the AT2 cells and was connected to a loss of AT2 cell identity during disease progression (Mayr, Simon et al. 2021). Moreover, CRTCA1 levels were negatively correlated with those of SOX4 whose transcription was shown by (Strunz, Simon et al. 2020) to regulate AT2 differentiation process.

As human and mouse scRNA-seq datasets begun to amass, the first large studies begun to emerge. Such a big project was undertaken by Huang and colleagues (2021) with the ultimate goal of disentangling the various ‘intermediate’ epithelial cell types previously described in human and mouse samples. More specifically, through the integration of multiple publicly-available scRNA-seq datasets, they demonstrated various differences between human and murine intermediate cell types that could be attributed to either murine model limitations or inter-species particular characteristics (Huang and Petretto 2021). In addition, IPF AT1 cells were more closely related to murine progenitors than to normal AT1 human cells, suggesting an incomplete differentiation as a common baseline of fibrotic lungs. Moreover, in contrast to the proposal of (Adams, Schupp et al. 2020, Habermann, Gutierrez et al. 2020) that senescence-related genes emerge only in latter stages of aberrant basaloid cells expansion, Huang and Petretto (2021) supported that it may actually play a more active role earlier in the process, similarly to PATS of (Kobayashi, Tata et al. 2020). On the other hand, SMAD3-activated TGF β signaling from committed KRT5-/KRT17+ cells is suggested to drive the cells’ pro-fibrotic phenotype in accordance to (Yao, Guan et al. 2020) findings. The authors also provide evidence that microinjuries in the IPF lung might sustain Notch-signaling in transitional AT2 cells, thus

inhibiting their homeostatic developmental course in favor of KRT5-/KRT17+ cells creation (Huang and Petretto 2021). Last, an attempt to order all AT2-AT1 intermediate cell states was made, with Cdc42-null AT2s (Wu, Yu et al. 2020) being the most remote from the rest of the cells and Krt8+ ADI (Strunz, Simon et al. 2020) and DAPT cells (Choi, Park et al. 2020) the most closely 'related'.

Impaired epithelial turnover seems to be a feature common between several lung pathologies. Indicatively, KRT5-/KRT17+ basaloid cells were detected in SSc-ILD samples along with a unique transcriptional profile of fibrotic lung AT1 cells further strengthened already existing data of a common pathogenic mechanism between lung fibrotic pathologies (Valenzi, Tabib et al. 2021). In contrast, non-UIP ILD and IPF alveolar basal cells have been reported to have distinct transcriptional profiles (Jaeger, Schupp et al. 2022). The latter had a KRT17^{hi}/PTEN^{low} phenotype characteristic of a de-differentiation status and were transcriptionally similar to the aberrant basaloid cells described previously (Adams, Schupp et al. 2020, Habermann, Gutierrez et al. 2020). Last, pro-fibrotic characteristics of IPF basal cells, such as bronchospheres formation, engraftment and proliferation capabilities, were attenuated by SRC-inhibition (Jaeger, Schupp et al. 2022).

Based on previous knowledge that mesenchymal cells do support epithelial cells physiological function in the lung, Xie and colleagues (2021) whether pro-fibrotic fibroblasts-produced factors are the reason for the impaired epithelium regeneration during PF. Motivated by the decreased expression of growth hormone receptor (GHR) in both human and mouse fibrotic lungs, which correlates with decreased lung functionality measurements, the authors supported that mesenchymal cells supply the lung epithelium with GHR (Xie, Kulur et al. 2021). Inhibition of this supply disrupts renewal of the latter leading to fibrosis establishment. Their data along with those of (Wu, Yu et al. 2020) shape an epithelium-mesenchyme pro-fibrotic signaling loop functionally connecting those two compartments during PF.

Last, CD274 (PD-L1) which is suggested to regulate fibroblast to myofibroblast trans-differentiation and collagen deposition, has been found to be specifically expressed by AT2 cells, aberrant basaloid and KRT5-/KRT17+ cells of the lung (Ahmadvand, Carraro et al. 2022). FACS enrichment of epithelial cells led to the discovery of a new CD274+ cell group expanded in the IPF tissues and expressing moderate levels of AT2 marker genes. From both a molecular and functional perspective, those cells were very similar to mouse injury-activated alveolar epithelial progenitors (IAAPs) (Ahmadvand, Lingampally et al. 2022) and demonstrated self-renewal capacity in the steady state lung (Ahmadvand, Carraro et al. 2022).

1.7.3.2 Single myeloid cells of the lung

Myeloid cells constitute another important player for the initiation and establishment of fibrosis and thus, have become the research subject of many single cell publications.

Initially, based on an RNA-seq analysis of FACS-enriched myeloid sub-types of the lung, it was suggested that monocyte-originating and not lung-resident macrophages are the ones implicated in fibrosis development (Misharin, Morales-Nebreda et al. 2017). Murine data from the same publication supported the existence of a continuous differentiation path between monocyte-derived and alveolar macrophages, with these two populations accounting for pro-fibrotic and homeostatic conditions, respectively.

Similarly to the case of epithelial cells, Reyfman and colleagues (2019) have identified macrophage clusters specifically enriched for cells of fibrotic origin. Pathway analysis of deregulated genes in the cluster yielded fibrosis-related processes, such as “extracellular matrix organization” and “regulation of migration” (Reyfman, Walter et al. 2018).

In support of a continuous periphery-to-lung developmental trajectory of myeloid cells, Aran and colleagues (2019) examined murine fibrotic niches and identified a population of transitioning macrophages (Aran, Looney et al. 2019). These SiglecF⁺CD11c⁺MHCII^{hi} cells were observed in contact with fibroblasts and regulated their proliferation via Pdgfaa secretion in both human and mouse lungs. Ablation of this macrophage population inhibited fibrosis by affecting fibroblast expansion and fibrotic scar formation (Aran, Looney et al. 2019). Further supporting evidence came from an independent publication where monocyte-derived alveolar macrophages of the fibrotic niches were found to be important sources of Pdgfa regulating fibroblasts pro-fibrotic behavior (Joshi, Watanabe et al. 2019). On the other hand, tissue-resident interstitial macrophages did not contribute to the niche-associated macrophage group. Importantly, (Xie, Wang et al. 2018) dataset re-analysis produced similar results.

PF heterogeneity is reflected, among others, between the various lung locations, with upper lobe parts presenting milder fibrosis than lower parts. To examining this difference, a single cell comparative analysis revealed an increase in fibroblast, basal and club cell numbers in lower lobe sections, in contrast to alveolar macrophages (Morse, Tabib et al. 2019). Clustering of the latter revealed three groups in both control and IPF lungs: FABP4^{hi}, SPP1^{hi} and FCN1^{hi}, each one with a unique enrichment pattern. More specifically, SPP1^{hi} macrophage numbers were increased in the lower fibrotic lung parts in comparison to the upper IPF and control lungs, while staining for SPP1 suggested localization at the fibrotic niches (Morse, Tabib et al. 2019). In addition, SPP1^{hi} cells of the control lungs were characterized by the expression of quiescent cell cycle markers, whereas they were found to be heavily cycling in all IPF sampling sites. Subsequent gene network analysis pinpointed the densest part of the network between fibroblasts and fibrotic niche macrophages and it was thus suggested that tissue resident macrophages that undergo changes during fibrosis are those that promote pathogenesis, contrarily to the findings of (Reyfman, Walter et al. 2018, Aran, Looney et al. 2019, Joshi, Watanabe et al. 2019).

The aforementioned findings were replicated by Valenzi and colleagues (2019) as macrophages and monocytes were separated into the same three groups described in (Morse, Tabib et al. 2019) and a SPP1^{hi} macrophages abundance was found increased in fibrotic samples compared to respective controls (Valenzi, Bulik et al. 2019). SPP1-marked macrophages were then re-identified when an archetype analysis led to the discovery of a pro-fibrotic SPP1⁺CSF-1⁺ subtype at the very end of the respective trajectory (Adams, Schupp et al. 2020). Last, the “activated alveolar macrophages” reported by Mayr and colleagues (2021) was an independent identification of a SPP1^{hi} macrophage population (Mayr, Simon et al. 2021).

Further interest in the PF myeloid compartment was triggered by (Fraser, Denney et al. 2021) who observed that peripheral macrophages CD64 expression correlates with patients’ fibrotic load. More specifically, it was shown that IPF-suffering individuals had an increased number of monocytes in the periphery accompanied by elevated levels of serum IL6, CCL-2 and CSF-1, a known promoter of epithelium senescence. Re-analysis of (Reyfman, Walter et al. 2018) single cell data identified, among others, a transitional macrophage type with increased levels

of CCL-2 and type-I IFN, suggesting a lung origin for this periphery-detected feature (Fraser, Denney et al. 2021). From a transcriptional point-of-view, this macrophage subset was similar to the pro-fibrotic group of (Adams, Schupp et al. 2020). Furthermore, a decrease in the numbers of mature macrophages in favor of expanded monocyte and transitional macrophage populations in the fibrotic lung corroborated the findings of previous publications such as that of (Aran, Looney et al. 2019).

Extending previous findings of their lab and extending the work of (Fraser, Denney et al. 2021), Valenzi and colleagues (2021) detected differential regulation of interferon signaling features between IPF and SSc-ILD macrophages. In more detail, IFN γ signaling was intensified in SSP1^{hi} and FABP4^{hi} macrophages of IPF versus those of SSc-ILD origin, while the opposite pattern was followed by IFN-I, thus necessitating a more careful examination of any IPF patients interferon treatment (Valenzi, Tabib et al. 2021).

1.7.3.3 Single cells of the fibrotic mesenchyme

Stromal cells and especially fibroblasts have long been in the crosshairs of scientists investigating lung fibrosis, as these cells promote tissue stiffness via direct ECM remodeling. Single cell analysis of this lung compartment has uncovered a previously unknown great heterogeneity and have also challenged the effectiveness of long established marker genes. Furthermore, the ever-increasing number single cell datasets attempt to reveal the source of pro-fibrotic fibroblast populations, yet with mostly conflicting results.

One of the most heavily researched IPF single cell topics is about the discovery and characterization of *in vivo* mesenchymal sub-types. During the early days of IPF single cell analyses, researchers observed the existence of two mesenchymal niches: an Axin2⁺ myofibrogenic and a Pdgfra⁺ Wnt-responding alveolar one (Zepp, Zacharias et al. 2017). Lineage tracing of single cells from a BLM murine model revealed that the former cells give rise to pathogenic myofibroblasts and were thus termed Axin2⁺ Myofibrogenic Progenitors (AMPs). The latter group responded to acute lung injury by providing the necessary support for the physiological AT2 regeneration and trans-differentiation into AT1s and were thus called mesenchymal alveolar niche cells (MANC) (Zepp, Zacharias et al. 2017). A year later, examination of FACS-enriched mesenchymal cells from both BLM-treated and untreated mice led to the identification of multiple fibroblast clusters, whose comparison led to several interesting findings (Xie, Wang et al. 2018). As far as cluster markers are concerned, *Hhip*, *Aspn* and *Mustn1* proved better myofibroblast markers compared to *Acta2*, as the latter was expressed indistinguishably by both myo- and matrix fibroblasts. Detected lipofibroblasts expressed *Adrp* and *Pparg* in addition to already known markers and were also characterized by an M2 macrophage-like signature blurring the line between the two cell types. In contrast to previous reports, neither lipofibroblasts nor myofibroblasts expressed elevated levels of *Pdgfra*, a feature which was primarily produced by *Coll3a1* and *Coll4a1* matrix fibroblasts. A rare *Pgfrb*⁺ population was found scattered across known fibroblast clusters and constituted predominantly by cells of fibrotic origin which did not overlap with the pericyte group. Juxtaposed to the observations of (Zepp, Zacharias et al. 2017) D0 myofibroblasts and D21 Pdgfrb^{hi} cells correlated well with AMPs, while MANCs were more closely related to the here described *Coll3a1* and *Coll4a1* matrix fibroblasts. Collectively, the authors suggested that the term ‘pericytes’ may actually describe a heterogeneous set of cells made of both myofibroblast and matrix fibroblasts and that *Pdgfra*⁺ cells are plastic matrix fibroblasts that can

transdifferentiate into lipofibroblasts. Last, it was proposed that the latter give rise to myofibroblast via $Pgfrb^{hi}$ intermediate mesenchymal cells (Xie, Wang et al. 2018).

In a different cell isolation strategy, sampling various lung regions led to the discovery of a $HAS1^{hi}$ ECM-producing mesenchymal population occurring primarily in the fibrotic lung (Habermann, Gutierrez et al. 2020). Located at peripheral and subpleural areas of the organ, $HAS1^{hi}$ cells were encountered in deeper tissue parts during PF, indicating a potential invasive phenotype. Their transcriptome was enriched with epithelial-to-mesenchymal transition and cellular stress processes. Lineage receptor analysis interconnected fibroblasts and epithelial cells via matrix-driven integrin-dependent signaling pathways (Habermann, Gutierrez et al. 2020). The above described cell types were recovered from a re-analysis of (Reyfman, Walter et al. 2018, Morse, Tabib et al. 2019) datasets and similar transcriptional profiles including elevated expression of *HAS1* have been independently reported by (Adams, Schupp et al. 2020).

Another active research topic revolves around mesenchymal-related signaling pathways. To validate previous data indicating a *Pdgfra* flow from $SiglecF+CD11c+MHCII^{hi}$ macrophages to fibroblasts located in adjacent fibrotic niches, (Aran, Looney et al. 2019) re-analyzed a murine lung single cell dataset. Although *Pdgfra* was not detected, Aran and colleagues identified a new group of cell cycling fibroblasts in the fibrotic samples. Ultimately, depletion of $CX3CR1+$ macrophages led to a significant reduction of cells expressing *Pdgfra* and *Pdgrb* fibroblast markers and a simultaneous reduction of collagen levels (Aran, Looney et al. 2019). Further strengthening the hypothesis of a macrophage-dependent fibroblasts expansion, single cell ligand-receptor analysis revealed that alveolar macrophages do target fibroblasts with pro-proliferating molecules, such as *Pdgfra* (Joshi, Watanabe et al. 2019). Moreover, analysis of a DEG network revealed an intense communication between fibroblasts and $SPP1^{hi}$ macrophages (Morse, Tabib et al. 2019).

With the aim of exploring early fibrosis processes, Peyser and colleagues (2019) collected murine lung cells on the 11th day of BLM treatment and examined their transcriptional profile. Interestingly enough, 49 genes could be found deregulated in an activated group of fibroblasts, but their expression was not limited to BLM-treated cells. This observation supported the notion that pro-fibrotic mesenchymal cells do emerge from homeostatic ones during fibrosis establishment (Peyser, MacDonnell et al. 2019). Furthermore, as many of these genes, including *Sfrp1* and *Cthrc1*, have been previously linked to anti-fibrotic processes, it was proposed that an increase in their transcription may be part of an active defensive mechanism. In terms of population size, no changes in fibroblast numbers were observed during the initial stages of the model, suggesting that mesenchymal cells and fibrosis initiation are not functionally related (Peyser, MacDonnell et al. 2019). Last, in accordance with (Xie, Wang et al. 2018) it was proposed that no single marker, let alone *ACTA2*, can be used for adequate myofibroblasts characterization.

Extending mesenchyme research, Valenzi and colleagues (2019) scrutinized the transcriptome of control and SSc-ILD cells and identified two major, $SPINT2^{hi}$ and $MFAP5^{hi}$, and one minor, $WIF1^{hi}$, group of fibroblasts present in both phenotypes. On the other hand, detected myofibroblasts were of fibrotic origin and intensively expressed cell proliferation-related features, such as *CTHRC1* and *ACTA2* (Valenzi, Bulik et al. 2019). Noteworthy, the latter was also transcribed by smooth muscle cells along with *MYH11* and *PLN*, further strengthening

the opinion that *ACTA2* use is inappropriate for myofibroblasts identification. Moreover, it was suggested that murine mesenchymal markers are not directly analogous to human cells. In agreement to (Peysers, MacDonnell et al. 2019), fibroblast numbers did not change during fibrosis, opposite to the proportion of myofibroblasts (Valenzi, Bulik et al. 2019). Last, in contrast to (Xie, Wang et al. 2018), it was proposed that myofibroblasts arise from a quiescent cell type, with MFAP5^{hi} group being the most probable developmental source (Valenzi, Bulik et al. 2019).

The latter hypothesis that pro-fibrotic mesenchymal cells arise from a resting steady state cell type was further reinforced by the work of (Adams, Schupp et al. 2020). More specifically, invasive fibroblasts were detected in parallel to myofibroblasts, with the latter forming aggregates in physical proximity to aberrant basaloid cells. Correlation between pseudotime distances and cells' transcriptome highlighted a slow and progressive increase in the expression of genes related to both aforementioned cell types.

A quite interesting observation about IPF mesenchymal progenitor cells (MPCs) is that their fibrogenic character seems to be imprinted and thus manifestable in the absence of a pro-fibrotic microenvironment. Transcriptome examination of IPF and control MPCs demonstrated that the biggest differences between phenotypes can be seen in the least differentiated cells (Beisang, Smith et al. 2020). The majority of those expressed CD44 and it was also shown that CD44^{hi} IPF MPCs produced more colonies outside of their native fibrotic milieu, in contrast to their CD44^{lo} IPF counterparts. From a spatial point-of-view, pathological MPCs were found at the high perimeter region of fibroblastic foci, similarly to many other fibrosis-related cell types. Collectively, data from this publication support the early emergence of pro-fibrotic IPF fibroblasts and the hypothesis of their emergence from macrophage-corrupted tissue resident control MPCs (Beisang, Smith et al. 2020).

Working on the hypothesis that fibroblasts achieve their diverse functional repertoire thanks to tissue-specific transcriptional profiles, Buechler and colleagues (2021) assessed an extended research area of multiple integrated single cell samples originating from various tissues, diseases and two distinct species. In precis, it was revealed that fibroblasts can be separated into a universal, a steady-state and an activated (pathology-related) group (Buechler, Pradhan et al. 2021). More importantly, researchers witnessed a high concordance between mouse and human fibroblasts, while multiple PF cells expressed various universal markers (Habermann, Gutierrez et al. 2020).

In search for ECM-supporting cells, Tsukui and colleagues (2020) identified Cthrc1+ and CTHRC1+ fibroblasts as the main producers of secreted collagen. In detail, normal mouse fibroblasts were separated into four alveolar subgroups, two types of adventitial and a cluster of bronchial fibroblasts (Tsukui, Sun et al. 2020), all in different locations similarly to (Habermann, Gutierrez et al. 2020). Comparison of their transcriptome with expression patterns from previous publications led to the identification of cells homologous to (Xie, Wang et al. 2018) fibroblasts and AMPs (Zepp, Zacharias et al. 2017). In contrast, lipofibroblast markers were detected across several cell groups (Tsukui, Sun et al. 2020). Interestingly, Acta2 has not only been once more accused as an unreliable myofibroblast marker, but it was also proposed as uncharacteristic of collagen producing cells in general (Tsukui, Sun et al. 2020). As far as the identified Cthrc1+ fibroblasts are concerned, they presented a stronger migratory potential than alveolar or adventitial fibroblasts and they also had an enhanced ability to

colonize BLM-treated murine lungs. From a developmental point-of-view, both RNA velocity and trajectory analysis indicated alveolar fibroblasts as the source of *Cthrc1*⁺ cells. Importantly, re-analysis of (Peyser, MacDonnell et al. 2019) data identified a latent fibrosis-specific cell cluster similar to the *Cthrc1*⁺ fibroblasts described. Similar findings emerged from human scRNA-seq data and re-analysis of (Morse, Tabib et al. 2019) dataset in which a collagen-producing *CTHRC1*+*ACTA1*^{low} population was identified. Last, *CTHRC1*+ fibroblasts were also described independently in scleroderma patients (Valenzi, Bulik et al. 2019).

Approaching pathogenic fibroblasts establishment from both a transcriptomics and proteomic point-of-view, Mayr and colleagues have observed that myofibroblast signatures are up-regulated early in the progression of the disease (Mayr, Simon et al. 2021), in contrast to the suggestions of (Peyser, MacDonnell et al. 2019). In addition, it was suggested that lung *PDGFRB*⁺ cells which are often vaguely labelled as perivascular pericytes, are more heterogeneous than previously appreciated, encompassing not only perivascular pericytes, but also fibroblasts and smooth muscle cells (Mayr, Simon et al. 2021). Similar results regarding the heterogeneity of *PDGFRB*⁺ cells were produced by (Wu, Yu et al. 2020).

In extension of the (Mayr, Simon et al. 2021) findings, Nakahara and colleagues detected meflin (*ISLR*), a gene up-regulated during fibrosis, in a population of *PDGFRB*⁺/*EPCAM*⁻/*CDH5*⁻/*PTPRC*⁻ cells. The latter most probably hosted fibroblasts and myofibroblasts of both the control and the IPF lung (Nakahara, Hashimoto et al. 2021). Examination of the spatial distribution of *ISLR*⁺ fibroblasts in control lungs revealed their intermittent presence in perivascular or periepithelial regions. During fibrosis three quarters of them were located in fibrotic lesions with half of them being *ACTA2*⁻, while *ACTA2*⁺ fibroblasts that were found in dense fibrotic regions were practically negative for *ISLR*. Data from further experimentations revealed that meflin is required for restraining the pro-fibrotic and senescent lung fibroblast phenotype (Nakahara, Hashimoto et al. 2021). Conclusively, it was proposed that in early fibrogenesis, meflin increased expression is inhibited by persistent *TGFβ* signaling, ultimately leading to fibroblast senescence and aberrant pro-fibrotic activity (Nakahara, Hashimoto et al. 2021).

It may be that mesenchymal cells are directly responsible for the secretion of ECM-remodeling factors, thus leading to progressive lung stiffness establishment, yet there are also other tissue compartments that can indirectly affect the whole process. Extending previous knowledge and data presented by (Habermann, Gutierrez et al. 2020), Xie and colleagues (2021) have for the first time proposed a fibroblast-epithelial communication line via *GHR*-enriched vesicles. More specifically, *GHR* was found under-expressed during fibrosis and it was revealed that its presence is necessary for the successful recovery of an injured lung epithelium (Xie, Kulur et al. 2021). Diminished exosomal and cellular *GHR* expression from mesenchymal cells suppress epithelial cell progenitor renewal, thus leading to an impaired epithelium recycling. Except for the signals exchanged between epithelial and mesenchymal lung cells, current data also propose an active signaling channel between fibroblasts and endothelial cells (Xiucheng, Xichun et al. 2021). Beginning from a clustering of fibroblast cells, authors were able to identify all common cell types, such as myo- and lipo- fibroblasts, while they were also in place to detect *Cthrc1*⁺ collagen producing fibroblasts as previously reported by (Tsukui, Sun et al. 2020). Ligand-receptor analysis between fibroblast and epithelial cells suggested a potentially important contribution of the epithelium in the regulation of fibroblast proliferation and ECM

production. Last, fibroblasts were witnessed to communicate with macrophages. The latter, expressing PLA2G7 during fibrosis were found to promote fibroblast-to-myofibroblast transition via the LPC/ATX/LPA/LPA2 axis (Wang, Jiang et al. 2022). Pharmacological inhibition of Pla2g7 ameliorated fibrosis in BLM-treated mice and reduced the expression of many pro-fibrotic features, including *Acta2* and *Col1a1*. On the contrary, expression of *ACTA2*, *FNI* and *COL1A1* increased in 3T3 cells when co-cultured with PLA2G7^{hi} macrophages compared to empty vector transfected ones.

As PF is an age-related disease, examination of human and mouse mesenchymal samples from various life stages can provide illuminating information for steady state and pathologic conditions. In such a research initiative (Liu, Rowan et al. 2021) scrutinized murine mesenchymal cells in multiple age time-points. To begin with, lipofibroblasts were located at the one end of E17.5 cells trajectory analysis opposite to fibroblasts and myofibroblasts. Subsequent examination of lipofibroblasts transcriptome across multiple developmental stages yielded a novel set of marker genes capable of identifying lipofibroblasts at all time-points studied. In contrast, the commonly-used gene set was capable of identifying the cell type only during its peak stage. Similar results were also obtained for human lipofibroblasts with TCF21 being the most reliable marker of the group irrespective of the developmental stage. As far as myofibroblasts are concerned, discriminative genes were identified for murine but not for human cells, as human myofibroblasts were inseparable from smooth muscle cells. Interestingly, on E14.5 researchers witnessed the emergence of a transcriptionally-unique Ebf1+ mesenchymal population. Following the signal into adult mouse lungs, Ebf1 signal was detected in pericytes and fibroblasts bearing the same embryonic expression signature, once again suggesting a common developmental origin for both the aforementioned cellular groups. Homologues cells were found in human IPF and control lung samples, while comparison of deregulated functions revealed the enrichment of various pro-fibrotic processes, such as fibroblast migration and proliferation, during fibrosis. Regarding cell proportions, not any change in myofibroblast numbers nor evidence of trans-differentiation from other mesenchymal cell types was witnessed in neither species. On the contrary, it was proposed that all mesenchymal cells over-express ECM-related features. Last, re-analysis of the (Xie, Wang et al. 2018) dataset revealed that *Col14a1* and *Col13a1* fibroblast markers from were expressed by lipofibroblasts and Ebf1+ cells, respectively.

1.8 Integration efforts and meta-analysis

Nowadays, publicly accessible repositories host an ever expanding collection of gene expression datasets as a result of the unceasing research of the PF transcriptional landscape. In parallel to the increasing data volume came several attempts of datasets integration, combination and/or meta-analysis, aiming to reveal latent yet potentially informative patterns of disease manifestation, establishment and progression. Due to the complex nature and the huge size of many of these data collections, machine learning (ML) applications have been adopted multiple times in an attempt to delimit otherwise unconceivable pathological motifs.

1.8.1 Challenges in data meta-analysis

Dataset meta-analysis and/or combination presents several data quality and data collection-related challenges. To begin with, both publication and search bias can greatly affect the ‘raw materials’ of a meta-analysis and in turn quality of its results (Walker, Hernandez et al. 2008). Publication bias refers to the current preference of the scientific world to ‘success story’ publications, a notion that excludes negative, unpublished yet highly informative research

deliverables. On the other hand, search bias concerns the fragile dependence of search engine responses from user-provided keywords, a relationship that if not taken into consideration can conceal significant publications. In addition, selection bias is another pitfall that arises when the final dataset shortlist is not shaped based on a comprehensive set of criteria. Last, persisting inconsistencies in public datasets recording can affect meta-analytic results. Thus, data harmonization, careful manual curation and meta-data gleaning is a necessary data pre-processing stage.

Apart from the target data *per se*, a comprehensive meta-analytic effort can also be confined by the quality and availability of meta-data and annotation sources. Annotating omics data with biological and technical information can be extremely challenging due to often outdated public archive entries and lack of interconnection between biological databases. Together with the aforementioned difficulties, retrieving annotation for the so called ‘legacy’ data frequently proves to be a nearly overwhelming task. To overcome the above obstacles, manual curation of annotation sources is usually required, thus introducing a time-consuming and rather baffling step in the meta-analytic process (Krassowski, Das et al. 2020). Moreover, data heterogeneity introduced not only across but also within omics modalities due to for example differences in protocols and sequencing platforms used, do create the necessity of data wrangling via various transformation steps, such as normalization and standardization (Krassowski, Das et al. 2020).

1.8.2 Meta-analytic/ integration efforts in IPF

Collective analysis of independent IPF-centered research deliverables begun rather early with one of the first such publications been based on literature reported results as were (Tzouveleakis, Harokopos et al. 2007). Nevertheless, the authors suggested a novel set of PF biomarkers and distinguished *HIF-1* as an important player for epithelium homeostasis. In a more statistically oriented attempt, Kim and colleagues (2015) integrated clinical, mRNA and miRNA expression data from a single cohort of lung disease patients. Applying a custom pipeline of feature concatenation and dimensionality reduction followed by unsupervised clustering of smoothed feature intensities, the authors retrieved three significant patient clusters representing COPD, ILD and intermediate cases, each one with specific deregulated features (Kim, Herazo-Maya et al. 2015).

Other meta-analytic attempts aimed to illuminate the regulatory mechanisms governing PF progression. Towards that goal, weighted gene co-expression network analysis (WGCNA) has been applied on differentially expressed features isolated from two of IPF and control sample sets, while the two independent datasets have been used for validation purposes (McDonough, Kaminski et al. 2019). Specifically, 16 expression modules have been shaped and nine of them correlated with lung function measurements. Totally, immune response was found down-regulated during IPF in favor of an up-regulated humoral response, in parallel to two simultaneous fibrotic processes. Last, transcriptional co-activators p300 and TCF12 along with miR-205 and miR-30s were proposed to interconnect with specific gene co-expression modules (McDonough, Kaminski et al. 2019).

Data integration has also been exploited for purposes of drug repurposing. In such an example, differential expression and network analysis created a multi-dataset-based list of deregulated features capable of describing IPF heterogeneity both across individuals and between different disease stages (Karatzas, Bourdakou et al. 2017). At least five different drug candidates were proposed for all or for the two examined IPF stages specifically. Two years later, connectivity

map analysis (CMap) of an aggregated feature list produced another set of eight potential drugs (Wang, Zhu et al. 2019).

Identification of latent yet important disease targets/biomarkers has greatly benefited from combinatorial dataset analyses. To begin with, WGCNA of between-datasets most variable genes shaped two co-expression modules significantly correlated with disease status (Wang, Zhu et al. 2019). Examining their members' distribution the authors identified 30 hub features supposedly suitable as IPF biomarkers. Towards the same goal, Li and colleagues (2020) created a PPI network from the DEGs shared between six lung microarray datasets comparing IPF and control individuals. 24 hub nodes were identified and subsequent pathway analysis (PA) implicated them in several ECM-related pro-fibrotic processes (Li, Liu et al. 2020). An almost identical approach based on just a couple of datasets was published the same year reporting ten disease targets (Xu, Mo et al. 2020). In a third network-based attempt, 20 hub genes were pinpointed as important in pulmonary fibrosis. Many of these features were also found to be potential miRNA interactors according to their opposite deregulation pattern during PF (Zhu, Xu et al. 2021). Last, yet another network-revolving publication proposed 11 IPF target genes following 7 transcriptomics datasets integration and wet laboratory validation of the *in silico* findings (Wan, Huang et al. 2021). The same number of significant genomic features was disclosed by (Cui, Ji et al. 2021)

Analyzing a bigger sample size increases the chances of revealing otherwise-masked disease subtypes. Agglomerative hierarchical clustering of IPF patients from multiple datasets according to their DEGs led to the characterization of two major sub-clusters (Kim, Jung et al. 2021). Each sub-cluster was potentially driven by different cell types and was enriched in a specific biomarker set developed from a PPI network analysis. Moreover, key driver analysis (KDA) of the network's primary component produced 119 key disease genes, $\frac{1}{4}$ of which were also differentially regulated.

To accelerate future combinatorial research efforts, Villasenor-Altamirano and colleagues (2019) developed an omics database hosting expression datasets from IPF, COPD and control cases. PulmonDB, as it is called, enables researchers to interactively investigate changes in pulmonary transcriptome during disease without the need of any computational expertise. Nevertheless, knowledge of the R programmatic language is necessary to retrieve data for any downstream analysis (Villaseñor-Altamirano, Moretto et al. 2020).

In conclusion, PF datasets meta-analyses aimed to address some of the most common yet unsolvable problems in the field of lung pathology. Future evaluation of their results will judge the utility of such methods.

1.8.3 Machine learning in IPF

ML is a computer science discipline that aims on solving practical problems by algorithm construction and tuning (Burkov 2019). It has evolved dramatically from the beginning of the 21st century and has nowadays various applications in a great spectrum of fields, including biology and related fields (Jordan and Mitchell 2015). ML trains mathematical formulas on a (non) labelled input dataset (training data) and then uses them to predict the labels of an unknown dataset. This type of 'learning' differs significantly from the conventional learning encountered in the animal kingdom, as it strongly depends on the quality and structure of the training data. Even the smallest divergence of the unseen data from the training standards will most probably negatively affect the performance of the model (Burkov 2019).

Distinguishable by the annotation level accompanying the training dataset, there are four types of ML, with supervised and unsupervised learning being the most famous ones. In contrast to the labelled examples provided in a supervised learning task, unsupervised learning is trained on a vector of unlabeled examples which is subsequently either transformed or used as is to provide an answer to a given task. For example, while supervised learning can be used for spam e-mail detection, unsupervised learning is often applied for clustering and dimensionality reduction purposes (Burkov 2019). Popular algorithmic examples developed for the latter application are uniform manifold approximation and projection (UMAP) (McInnes, John et al. 2020), autoencoders (Rumelhart, Hinton et al. 1986) and principal component analysis (PCA) (Jolliffe and Cadima 2016), one of the very first methods of this kind. Although multiple concerns are now beginning to arise (Dyer and Kording 2023), PCA and UMAP are still the methods of choice for transcriptomics data analysis bearing various advantages, such as those described in (Becht, McInnes et al. 2019).

In biomedical sciences, ML algorithms have been repeatedly used to analyze complex data and reveal previously unnoticeable motifs. For example, novel biomarkers have been proposed via ML for cancer (Huang, Cai et al. 2018) and NAFLD cases (Han, He et al. 2022). Another popular application is that of endotype discovery and differential response to treatment prediction, for example in acute respiratory syndrome (Calfee, Delucchi et al. 2018).

Concerning IPF, there are multiple research deliverables making use of ML for various reasons. To begin with, IPF diagnosis is a difficult task impeded by the similarities of its radiology and histopathology patterns to those of other ILDs. Currently, presence of a usual interstitial pneumonia (UIP) pattern is a necessary factor for successful IPF recognition. Although, high resolution computed tomography (CT) can ascertain UIP, its definitive existence can only be validated via examination of biopsy samples (Lynch, Saggarr et al. 2006, Raghu, Collard et al. 2011). In an attempt to avoid such an invasive method, Kim and colleagues (2015) used both microarray and RNA-seq gene expression data from surgical biopsy samples in order to develop a classifier algorithm capable of predicting UIP. Top 200 differentially expressed features between UIP and non-UIP samples were used in different size combinations to train a per technology classifier (Kim, Diggans et al. 2015). For microarrays a support vector machine and a lasso classifier were assessed, while for RNA-seq a linear support vector machine was applied on log transformed normalized data. Both classifiers achieved a >90% specificity, while a sensitivity of 82% (%CI 64-95) and 59% (35–82) was achieved by the microarray and the RNA-seq-based model, respectively. One more such attempt was based on gene expression data from trans-bronchial biopsies characterized or not by a UIP pattern (Pankratz, Choi et al. 2017). Quality controlled RNA-seq data were subjected to feature selection via logistic regression with elastic net penalty and a signature of 169 genes was used to classify patients. Ultimately, the aforementioned deliverables provided evidence that UIP prediction is feasible by examining biopsy transcriptional data, paving the way for the search of more similar signatures.

To rigorously examine their Envisia genomic classifier, Choi and colleagues (2018) provided further data supporting that it remains practically unaffected from unfavorable data characteristics, such as small sample sizes and heterogeneous pathology composition of the non-UIP samples (Choi, Liu et al. 2018). The model was also tested for its robustness against various quality and biological contamination levels using count data from 190 genes collected from various total mRNA datasets (Choi, Lu et al. 2017). It was revealed that Envisia can

adequately perform with UIP and non-UIP signals skewed up to 20% and 60%, respectively, while blood contamination can be tolerated up to 22%. Moreover, the test was found to be reproducible both within and across laboratories (Choi, Lu et al. 2017).

To further scrutinize their algorithm, Raghu and colleagues (2018) have trained Envisia on a new set of expression data and compared the predicted diagnosis to that of high resolution CT scans. Interestingly, Envisia was witnessed to objectively distinguish UIP from non-UIP even in cases of uncertain or inconclusive high-resolution CT scans (Raghu, Flaherty et al. 2019). Last, the genomic classifier was shown to increase the diagnostic certainty when used together with bronchoscopic lung biopsies and can even be used on its own given the appropriate diagnostic settings (Kheir, Alkhatib et al. 2020).

ML has also been exploited for disease severity recognition and classification. Towards that goal, unsupervised clustering of IPF/UIP and control samples based on expression data from a single microarray dataset, revealed six distinct subgroups of patients that correlated with disease severity and physiological lung function (Wang, Yella et al. 2017). Molecular signatures able to distinguish IPF from control and mild from severe IPF cases were also reported. Findings were validated in three independent microarray datasets. In addition, differential expression analysis (DEA) between the six subgroups produced approximately 3000 deregulated genes organized into three gene expression modules, each one enriched in different functional processes (Wang, Yella et al. 2017). Interestingly, DEA between patient groups identified both group-specific and universal features with the vast majority of the latter being deregulated in all three validation cohorts. A logistic regression classifier trained on the main dataset and tested on each of the validation ones, achieved specificity, sensitivity and accuracy >90% in all testing sets and proposed the universal set of DEGs as a robust IPF versus control classification signature (Wang, Yella et al. 2017). Last, classifier's training and validation with the unique gene set of the 6th subgroup enabled distinction of advanced from stable IPF cases.

On top of expression-based ML algorithms there are others that use imaging data in order to address, for example, limitations in manual or semi-automatic steps of high-resolution CT scans quantitative analysis (Barnes, Humphries et al. 2023). Indicatively, CT scan-trained neural networks have been developed to facilitate disease classification based on the criteria published by (Raghu, Collard et al. 2011) and (Lynch, Sverzellati et al. 2018). Ultimately, the algorithm performed at least the same with the majority opinion of a great number of specialized thoracic radiologists, thus providing a reproducible and quick way of patient stratification (Walsh, Calandriello et al. 2018). Furthermore, a novel deep convolutional neural network architecture strengthened by two layers of transfer learning has been developed for the classification of ILD patterns based on CT images (Huang, Lee et al. 2020). Last, unsupervised ML evaluation of CT images produced a set of radiological IPF progression markers capable of predicting disease outcome (Pan, Hofmanninger et al. 2023).

ML has also been employed for the discovery of disease biomarkers. Representatively, unsupervised separation of IPF subtypes led to the functional ranking of two IPF gene signatures and the isolation of a small group of BAL fluid markers separating disease from steady state (Wang, Yella et al. 2017). In another research deliverable, a random forest algorithm based on six expression datasets distinguished six genes, *CDH3*, *DIO2*, *ADAMTS14*, *HS6ST2*, *IL13RA2* and *IGFL2*, as the best IPF versus control biomarkers (Li, Wang et al. 2023).

Interestingly, all the expression of all the above features correlated with lung function, while that of *CDH3*, *DIO2* was also associated with patients' survival.

To summarize, ML can successfully address several yet unanswered biological questions and unsolved tasks. Novel algorithms development is promising to better model biological data and thus speed up biomedical research.

2. Materials & Methods

2.1 Transcriptomics data selection for Fibromine

Scientific literature and research papers were scrutinized in order to form a list of IPF-related transcriptomics datasets. Those with less than three biological replicates per condition were filtered out along with those not providing editable raw data. Focus was given on commercially-available transcriptomics technologies for reasons of annotation availability. Datasets addressing only homeostatic or pathologic samples were filtered out, too. The vast majority of transcriptomics datasets used can be found in GEO repository (Barrett, Wilhite et al. 2013). Each archived dataset is represented by a unique GSE identifier which comprehensively codifies the individual samples, the technology used and any recorded metadata (Barrett, Suzek et al. 2005).

Data of the aforementioned datasets were collected using *ad hoc* created packages. Specifically, GEOquery R package served as a GEO-proxy so as to easily fetch raw microarray data, as well as relative technical and biological annotation (Davis and Meltzer 2007). RNA-seq FASTQ files were retrieved from Sequence Read Archive (SRA) (Kodama, Shumway et al. 2012) as SRA files using the SRA Toolkit function *prefetch* and were then transformed with the *fastq-dump* utility.

2.2 Fibromine-hosted microarray data analysis

2.2.1 Data preprocessing

Microarray platforms can be separated into one or two color arrays, based on the number of color dyes used. Each array type requires different handling with one color data usually processed in three steps: background correction, normalization and summarization. On the other hand, two color data often require an extra specific step of within array normalization.

Affymetrix 3' IVT expression arrays, as well as their gene and exon successor platforms were consistently pre-processed using Robust Multichips Average (RMA), a procedure that has been repeatedly reported to yield superior results during downstream DEA (Cope, Irizarry et al. 2004, Harr and Schlötterer 2006). RMA can be summarized in three steps: background adjustment, quantile normalization and summarization of the expression values. Background adjustment deals with the noise arising from several sources, such as non-specific hybridizations. During RMA probes' GC content can be taken or not into consideration via algorithms implementation in the *gcrma* (Wu, Irizarry et al. 2004) and *rma* (Irizarry, Hobbs et al. 2003) package, respectively. *gcrma* was only applied to 3' IVT arrays as gene and/or exon platforms do not use mismatch (MM) probes. For the rest of the cases, RMA was used as implemented in the *affy* (Gautier, Cope et al. 2004) and *oligo* (Carvalho and Irizarry 2010) R packages. For gene/exon arrays, RMA was applied on the 'core' array probes. HTA arrays were analyzed the same way with the gene/exon ones.

Agilent microarrays can be divided into one or two channel arrays according to the number of colors used to measure gene expression. Data pre-processing requires within (only for two

color chips) and between arrays normalization post to background correction. *normexp* background correction with a 50 units offset was applied. This method was developed based on the RMA model and has been reported to perform better than other respective methods (Ritchie, Silver et al. 2007). The empirical offset value of 50 was added as a variance stabilization strategy necessary for the method's successful results. Within array normalization deals with the inherent dye bias of two color arrays. Global loess method was selected instead of the otherwise better print-tip loess, as Agilent technology does not use print-tip groups (Smyth and Speed 2003). Last, between arrays normalization was used to deal with systematic biases of both one and two color microarrays. To maintain the analysis as consistent as possible, quantile normalization methods extending from the homonym Affymetrix strategy were selected, being suited for both one and two color arrays (Yang and Thorne 2003). All the aforementioned processes were used as implemented in *limma* R package (Ritchie, Silver et al. 2007). Duke Operon arrays were analyzed as the Agilent two color ones.

Illumina BeadChip microarrays were pre-processed with the *neqc* method that combines *normexp* and background correction followed by quantile normalization, as it is known to have top performance for Illumina platforms (Shi, Oshlack et al. 2010).

2.2.2 Quality control, curation and probe filtering

Quality control for Fibromine microarray data was performed using the *arrayQualityMetrics* R package (Kauffmann, Gentleman et al. 2009) and *ad hoc* created code scripts on background corrected and normalized expression values. Identification and removal of outlier samples was performed in a recursive manner. The process was facilitated by diagnostic plots and scores hosted in a per-dataset *arrayQualityMetrics* html report. Outlier detection was further supported by independently-created top DEG-expression-based sample hierarchical clustering and sample location in a PCA reduced dimension space (Jolliffe and Cadima 2016). Datasets with poor quality and/or per-condition inseparable samples were discarded.

Subsequently, probes were filtered in order to increase the power of downstream DEA by decreasing the number of statistical tests made. Probes removal was based on biological (e.g. matching to none or multiple genes) and/or technical criteria (e.g. control and low variance probes) (Cordero, Botta et al. 2007). More specifically, control probes along with probes matching to no or multiple genes were filtered out. Respective annotation was found in either dedicated Bioconductor databases or array's manufacturer website. R packages such as GEOmetadb were also utilized (Zhu, Davis et al. 2008).

2.2.3 Gene level summarization and differential expression analysis

Post to probe selection, intensities were summarized at the gene level represented by HGNC gene symbols. Intensities were recorded *as is* in cases of 1:1 probe-gene matching, while the weighted average was taken for genes represented by multiple probes. *weighted.mean* function of the *stats* R package was used with weights summing up to the unit.

Statistical comparison of gene expression levels between conditions was performed using the *limma* moderated t-test statistics method (Ritchie, Phipson et al. 2015). Various experimental designs described in the package's vignette were applied wherever necessary and occasional batch effects were incorporated into the statistical model. Multiple testing bias was corrected using the false discovery rate (FDR) method (Benjamini and Hochberg 1995). Significantly deregulated genes were defined by an absolute fold change (FC) ≥ 1.2 and an FDR-corrected p-value < 0.05 .

2.3 Other microarray data analyses

2.3.1 Kidney disease datasets

GSE66494 (Agilent) and GSE104066 (Affymetrix) raw data were retrieved from GEO via *GEOquery* R package (Davis and Meltzer 2007). The Agilent dataset was background corrected with the *normexp* method and an offset of 50 units (Ritchie, Silver et al. 2007), prior to between arrays quantile normalization. Affymetrix data were background corrected and RMA normalized utilizing the *oligo* R package (Carvalho and Irizarry 2010). *arrayQualityMetrics* R package facilitated quality-control and outlier removal for both datasets; GSM1623315 and GSM3904846, GSM2788881 outlier samples were removed from GSE66494 and GSE104066, respectively.

Subsequent probe filtering of both datasets removed control, as well as probes matching to no or multiple HGNC gene symbols. Probes with intensity values close to the background as defined by interquartile range (IQR) calculation were also filtered out. Affymetrix probes with increased chances of cross-linking, as described in the technical annotation retrieved from *getNetAffx* function of the *oligo* package were removed, too.

Intensity values of both collections underwent gene level summarization with one:many gene:probe relationships represented by a weighted average (see section 2.2). Deregulated features were identified via the empirical Bayes test statistic implemented in *limma*. Absolute FC ≥ 1.2 and FDR corrected p-value < 0.05 were utilized to designate statistically significant expression changes.

2.4 Fibromine-hosted RNA-seq data analysis

For a consistent re-analysis, the vast majority of RNA-seq data included in Fibromine were fetched in a FASTQ file format from SRA (Kodama, Shumway et al. 2012) (see section 2.1). Reads alignment was performed in a two-step process coupling HISAT-2 (Kim, Paggi et al. 2019) and Bowtie2 (Langmead and Salzberg 2012) aligners. Samples as well as datasets with low alignment rate were filtered out. All downstream processes were completed using *metaseqR2*, an R package for streamlined RNA-seq handling, improved DEA and comprehensive results reporting (Fanidis and Moulos 2020). Outlier samples were identified and removed in a recursive manner via multi-dimensional scaling (MDS) plots and top DEG-based hierarchical clustering. MDS plots were created based on EDASeq normalized expression values.

2.4.1 Reads alignment

All RNA-seq datasets were mapped to GRCh38.p13 and GRCm38 reference genomes for *H. sapiens* and *M. musculus* samples, respectively. Use of the same per-species reference genome eliminated any chances of different annotation. Mapping was achieved via a pipeline of HISAT-2 and Bowtie2 aligners.

HISAT-2 (Hierarchical Indexing for Spliced Alignment of Transcripts 2) is a splice-aware aligner based on a graph Ferragina-Manzini index. It operates on a linear graph of the reference genome with alternative paths representing several genomic variations (Kim, Paggi et al. 2019). The aligner was applied on raw sequencing data without any modification of its default operating parameters. Exception to the rule is the *--rna-strandness* argument which was activated for the processing of paired-end stranded datasets. Reads that remained un-mapped were isolated for downstream processing.

Bowtie2 is a short read aligner permitting gapped mapping based on the Bowtie indexing approach (Langmead and Salzberg 2012). Reads that failed to map via HISAT-2 were fed to this aligner using the *-local* and *--very-sensitive-local* parameters. The first authorized the aligner to trim read ends if by doing so maximized the alignment score. The second matched to a preset of options that carefully fine tune the accuracy, sensitivity and speed of the local alignment. These options are: the allowance of up to 20 consecutive seed extension attempts, three-times re-seeding of reads with repetitive seeds, zero permitted mismatches, utilization of 20 nucleotide long seed substrings and usage of a gap between seed substrings given by $1 + 0.5 \times \sqrt{x}$ with x being the length of each read. In addition to the aforementioned two, the *--dovetail* argument was used for the alignment of paired-end reads in order for the mapping of dovetail pairs to be considered as concordant. Dovetailing describes the situation where one read mate aligns past the beginning of the other leading the wrong mate to begin upstream.

The produced SAM files were transformed into binary ones (BAM files) using *samtools view* function (*samtools* v1.7). BAM files from both aligners were then merged, coordinate sorted and indexed via *samtools merge*, *sort* and *index* functions, respectively. Datasets with very low proportions of overall aligned reads were discarded.

2.4.2 Reads quantification, normalization and filtering

Following alignment, BAM files were fed to *metaseqR2* package for downstream processing. Initially, read counting was performed via internal use of the *GenomicRanges* R package (Lawrence, Huber et al. 2013) and was based on the same reference genomes used during alignment. *count.type* argument of *metaseqR2* function was set to either 'gene' or 'exon' depending on each sequencing library.

Subsequently, in cases of exon counting, reads were filtered by default to ensure a minimum read presence. Thus, to maintain features with up to five exons, reads should have been mapped in at least two of them. On the other hand, more complex features were filtered in, if reads accumulated to at least 0.2 of their exons. Exon counts were then summarized to the gene level (*trans.level = 'gene'*).

Next, gene expression values were normalized prior to gene-level filtering. EDASeq two step normalization strategy was selected in order to address both gene length/ GC-content bias and library size/ composition (Risso, Schwartz et al. 2011). Before GC-correction, features with zero counts in total were automatically removed. As far as gene-level filtering is concerned, genomic features were filtered based on their structure, reads count, expression, biotype and presence. Specifically, genes shorter than 500kb or belonging to rare biotypes, such as *unprocessed pseudogene*, were filtered out. Genes with, on average, less reads per 100kb than the 25th quantile of the average count distribution per 100 base pairs were removed. Average count distribution was calculated based on each feature length. In cases of an exon count, all exons individual length was summed. Quantile was calculated per sample. Genes below the median of the overall count distribution were also filtered out. Last, the presence of a gene was evaluated across all samples, with genomic elements having less than 10 counts in the 25% of the samples being removed.

2.4.3 Statistical analysis

Normalized and pre-processed gene counts data were analyzed statistically for purposes of DEA. *metaseqR2* supports nine individual statistical methods including the popular DESeq2 (Love, Huber et al. 2014), edgeR (McCarthy, Chen et al. 2012) and limma-voom (Law, Chen

et al. 2014). Furthermore, it provides seven p-value combination methods that can be activated when multiple individual algorithms are selected. Herein, all nine baseline methods were calculated and then PANDORA (Moulos and Hatzis 2015) weighted p-value algorithm combined their results. Differentially regulated features were defined as those having an absolute FC ≥ 1.2 and a PANDORA meta-p-value < 0.05 . PANDORA p-values were preferred as they have been observed to more truthfully identify deregulated features. Briefly, PANDORA improves false hits control while maintaining true positives, while at the same time is robust against several RNA-seq data particular characteristics and biases (Moulos and Hatzis 2015, Fanidis and Moulos 2020).

2.5 Other RNA-seq data analyses

2.5.1 Inflammatory diseases data

metaseqR2 (Fanidis and Moulos 2020) was used to re-analyze expression data from immune-related diseases (Ota, Nagafuchi et al. 2021) for the purposes of (Nikitopoulou, Fanidis et al. 2021). EDASeq normalization was followed by default filtering and statistical analysis using DESeq (Anders and Huber 2010), DESeq2 (Love, Huber et al. 2014), limma-voom (Law, Chen et al. 2014), edgeR (McCarthy, Chen et al. 2012) and ABSSeq (Yang, Rosenstiel et al. 2016) individual algorithms. PANDORA (Moulos and Hatzis 2015) was then used to combine the produced p-values. DEGs were identified using an absolute FC ≥ 1.2 and a meta p-value < 0.05 .

2.5.2 Quant-Seq data

Quant-Seq (Lexogen) data (Moll, Ante et al. 2014) were analyzed in (Barbayaanni, Kanellopoulou et al. 2023). Initially, low quality read ends were trimmed (Phred < 20) by Trim Galore (<https://github.com/FelixKrueger/TrimGalore>). Then pre-processed reads were mapped against GRCm38 Ensembl genome in a two-step pipeline utilizing HISAT-2 (Kim, Paggi et al. 2019) and Bowtie2 (Langmead and Salzberg 2012) as described in section 2.4.1.

Resulting BAM files were downstream processed with *metaseqR2* (Fanidis and Moulos 2020). A 3' UTR reads count table was crafted with internal use of *GenomicRanges* (Lawrence, Huber et al. 2013). During counting, 3' UTR regions spanning minimum 300 base pairs, as well as 50 base pairs flanks were taken into consideration. Gene summarized reads were then EDASeq normalized (Risso, Schwartz et al. 2011) prior to default filtering. DEA used all nine individual methods provided and then PANDORA combined their p-values. FDR-corrected meta-p-values < 0.05 and absolute FC ≥ 1.2 were set as significance thresholds. Heatmaps were created using normalized and standardized expression values. Euclidean distance was used for hierarchical clustering.

2.6 TCGA data analysis

TCGA data were used for publications (Panagopoulou, Fanidis et al. 2021, Panagopoulou, Drosouni et al. 2022). In (Panagopoulou, Fanidis et al. 2021), TCGA database was queried for methylation and expression data of prostate adenocarcinoma (PC), lung adenocarcinoma (LC) and liver hepatocellular carcinoma (HCC) cases. Level 3 methylation (Infinium Human Methylation 450K bead-chip) and normalized RNA-seq data at the gene and isoform levels (Illumina HiSeq) were retrieved exploiting the *TCGAbiolinks* R package (Colaprico, Silva et al. 2016). Sample metadata accompanying were also downloaded. All available cases of matched tumor and tumor-adjacent control areas (15 for HCC, 35 for PC and 42 for LC) were selected in addition to 200 tumor samples per cancer. In the infrequent event of a single patient being represented by multiple files per modality, the weighted average of all available values

was used for subsequent processes; weights summed up to the unit. Relationship of differential expression and methylation events was assessed via Spearman's correlation test as implemented in the *cor.test* vanilla R function. FDR corrected p-value < 0.05 was set as statistical significance threshold. Small correlation coefficients were further examined by linear model fitting using the *lm* vanilla R function.

In (Panagopoulou, Drosouni et al. 2022), *TCGAbiolinks* was used to retrieve and process raw RNA-seq and level 3 methylation data, both originating from breast cancer patients. Sequencing data were normalized with EDASeq (Risso, Schwartz et al. 2011) and quantile filtered prior to DEA with *edgeR* (McCarthy, Chen et al. 2012). Absolute FC value ≥ 1.2 and adjusted p-value ≤ 0.05 were set to distinguish DEGs. Proper phenotype representation by RNA and methylation samples was examined by multi-dimensional scaling (MDS) and principal component analysis (PCA), respectively. Spearman correlation was used to identify significant relationships between gene expression and DNA methylation ($\rho \geq 0.4$ and adjusted p-value ≤ 0.05).

2.7 Microbiomics data analysis

16S rRNA sequencing data targeting seven hot spot regions (V2-V4 and V6-V9) were analyzed for the purposes of (Galaris, Fanidis et al. 2022). DADA2 (Callahan, McMurdie et al. 2016) was used to process 16S rRNA FASTQ files. More specifically, 14bp were trimmed from the left end of each read as proposed by IonTorrent. Reads of at least 50bp and with less than four “expected errors” were kept for downstream analysis. Expected errors (EE) were computed based on nominal quality scores (Q) as thus:

$$EE = \text{sum} \left(10^{-Q/10} \right)$$

Contaminant sequences were detected by aligning pre-processed reads against the human and mouse genomes using the FastQ Screen tool (Wingett and Andrews 2018). Mammal sequences were removed from the raw FASTQ files which were then merged per tissue and diet prior to filtering and trimming repeat. Subsequently, DADA2 functions were used for denoising (*learnErrors*) and chimera removal (*removeChimeraDenovo*). Reads were then assigned to amplicon sequence variants (ASVs) that were in turn matched to specific taxa using SILVA database (Quast, Pruesse et al. 2013). To adjust for differences in gene copy numbers (GCN), abundance values were divided by the species-specific number of 16S gene copies as retrieved from rrnDB (v5.7 NCBI) (Stoddard, Smith et al. 2015). Taxa with no entries in rrnDB were removed from all downstream analyses.

Bacterial diversity was assessed by the metrics of microbial richness (observed number of ASVs) and biodiversity (Shannon's index) (Shannon 1948). β -diversity was compared between conditions using the Aitchison distance as it takes into consideration the compositional character of 16S rRNA sequencing data (Gloor, Macklaim et al. 2017).

2.8 scRNA-seq data analyses

Seurat R package (Stuart, Butler et al. 2019) was the workhorse of scRNA-seq data mining, re-analysis and results visualizations.

2.8.1 GSE122960 data

For the purposes of (Zannikou, Barbayianni et al. 2021), raw scRNA-seq data were fetched from GEO and re-analyzed similarly to the original publication (Reyfman, Walter et al. 2018). Cells from each sample were filtered according to their proportion of mitochondrial reads and

the number of detected features. The remaining reads were normalized with the *LogNormalize* method and a scale factor of 10k. Subsequently, PCA was applied on the scaled values of the top 3000 highly variable genes (HVG) and the most informative principal components were selected using heatmaps and Elbow plots. K-closest neighbors of each cell were then identified and shared nearest neighbor (SNN) graph was constructed. Louvain original algorithm, as implemented in *FindClusters* function, was exploited for modularity optimization and cluster identification. Consistently to the published analysis the t-distributed stochastic neighbor embedding (tSNE) non-linear dimensionality reduction method (Van Der Maaten and Hinton 2008) was selected for 2D depiction of cells. Cell typing was performed using known marker genes. Cell filtering parameters, selected principal components, clustering resolution and cell typing decisions of the re-analysis were dataset-specific and in accordance to the annotation of the original publication.

Once individually analyzed, samples were integrated per condition and then across-conditions using the Seurat anchor-based strategy. Briefly, a diagonalized canonical correlation analysis (CCA) was used to jointly reduce dimensionality of datasets prior to CCA vectors L2-normalization and mutual nearest neighbors (MNN) identification. The occurring cell pairs, aka anchors, were hypothesized to represent an across-dataset shared biological state and thus form the basis of the integration process downstream (Stuart, Butler et al. 2019). During integration, each cell received a correction vector representing a batch average in order to prioritize cells based on both their biological state and robust anchor correspondence. The process was then extended to multiple datasets using a tree alignment-like method. Integrated data were scaled, clustered and visualized as above. Cell type assignment of cells performed during individual sample analysis was fine-tuned whenever necessary so as to support a final round of cell typing based on known cell population markers. The Wilcoxon rank-sum test implemented in *FindMarkers* Seurat function was exploited for the identification of DEGs based on the “RNA” slot expression values. Absolute FC value ≥ 1.2 and Bonferroni corrected p-value < 0.05 were set as thresholds of differential features transcription. Macrophages and fibroblasts specific sub-clustering was performed with the aforementioned methods.

The aforementioned re-analyzed dataset was also mined for the evaluation of *SH3PXD2A* expression at the single cell level as presented in (Barbayaanni, Kanellopoulou et al. 2023). *SH3PXD2A*⁺ fibroblasts were defined by non-zero expression of the gene based on normalized counts. Differential expression analysis was applied as above.

2.8.2 COVID-19 data

Multiple publicly available scRNA-seq R and Python objects were mined for *ENNP2* expression during SARS-CoV2 infection and steady state (Nikitopoulou, Fanidis et al. 2021) using Seurat (Stuart, Butler et al. 2019). Python packages were processed in R after transformation with the SeuratDisk R package (<https://github.com/mojaveazure/seurat-disk>). Gene expression deregulation was assessed by Wilcoxon rank-sum test, with absolute FC ≥ 1.2 and Bonferroni-adjusted p-value < 0.05 selected as significant deregulation thresholds.

For the identification of lung plasmacytoid dendritic cells (pDCs) in (Bharat, Querrey et al. 2020) dataset, dendritic cells were isolated and subjected to PCA based on scaled expression values (*ScaleData*) of their HVG (*FindVariableFeatures*). An SNN graph was constructed sourcing data from the 30 first principal components (*FindNeighbors*) and clusters were

subsequently shaped with a 0.8 resolution (*FindClusters*). Cell typing was elicited from markers of the (Travaglini, Nabhan et al. 2020) cell atlas.

2.8.3 Kidney disease data

Kidney scRNA-seq data were scrutinized for the purposes of (Magkrioti, Antonopoulou et al. 2022). The respective dataset was fetched from <https://doi.org/10.5281/zenodo.4059315> and mined using Seurat (Stuart, Butler et al. 2019). With an initial pre-processing stage already in place, CD10⁺ and CD10⁻ objects were first normalized and then integrated using Seurat v3 anchor-based method (see section 2.8.1) and a set of 2000 HVGs. Cell typing and metadata were maintained as in the original publication (Kuppe, Ibrahim et al. 2021). Wilcoxon-rank sum test (*FindMarkers*) was used for all differential expression tasks with default parameters. Absolute FC ≥ 1.2 and multiple testing-corrected p-value < 0.05 were selected as DEA significance thresholds. Communication networks between CD10⁺/CD10⁻ proximal tubule and the rest of the cells grouped per population were predicted by SingleCellSignalR (Cabello-Aguilar, Alame et al. 2020). Ligand-receptor pairs of interest were retrieved from CellTalkDB (Shao, Liao et al. 2021) and added to the default SingleCellSignalR database. Circus plots were shaped by the *circlize* R package (Gu, Gu et al. 2014).

2.8.4 Fibromine-hosted data

Already analyzed scRNA-seq data of (Mayr, Simon et al. 2021) were downloaded as described in the *README* page of https://github.com/theislab/2020_Mayr. Raw counts from the *integrated_human_dataset.h5ad* object were loaded into R and normalized with the *LogNormalize* Seurat method and default parameters (Stuart, Butler et al. 2019). Prior to downstream analysis “empty” barcodes were removed. DEA was performed per cell type and between phenotypes in a pairwise-fashion using default parameters of *FindMarkers* Seurat function.

2.8.5 LCN2 expression assessment

Cellular expression of *LCN2* was explored in four respiratory system-oriented scRNA-seq datasets (Galaris, Fanidis et al. 2023). These datasets were found at the GEO entries GSE136831 (Adams, Schupp et al. 2020), GSE135893 (GSE135893_ILD_annotated_fullsize.rds.gz) (Habermann, Gutierrez et al. 2020) and the github repositories https://github.com/theislab/2019_Strunz (Strunz, Simon et al. 2020) and https://github.com/theislab/2020_Mayr (Mayr, Simon et al. 2021). All data processes were performed using Seurat R package (Stuart, Butler et al. 2019). The Wilcoxon-Rank sum test as implemented in *FindMarkers* function was utilized for deregulated features identification (absolute FC ≥ 1.2 ; Bonferroni-adjusted p-value < 0.05).

Filtered data from GSE136831 were normalized with a 10k scaling factor (*NormalizeData*), prior to HVGs identification (*FindVariableFeatures*) and scaling (*ScaleData*). Data were then recuded in space using PCA, and the 7 first principal components selected via the median of all findPC methods (Zhuang, Wang et al. 2022) were exploited for the creation of a closest neighborhood graph (*FindNeighbors*). Louvain clustering with a resolution of 1.3 was applied for clusters identification. Cell typing provided along with the object were adopted. Non-linear reduction of dimensions was achieved via UMAP (*RunUMAP*).

IPF and control originating cells were isolated from both the GSE135893 and the (Mayr, Simon et al. 2021) object. Their Rrw reads were log normalized (*NormalizeData*; default parameters) before further data mining. Barcodes with an assigned “empty” and “NA” or “Low-Quality

Cells” cell type were removed from (Mayr, Simon et al. 2021) and (Strunz, Simon et al. 2020), respectively.

2.9 Weighted Gene Co-expression Network Analysis

WGCNA (Langfelder and Horvath 2008) was utilized for the identification of gene modules significantly related to PF, with the ultimate goal of biomarkers identification. Two different co-expression analyses were performed for human (GSE10667, GSE24206, GSE48149, GSE47460_GPL6480, GSE53845, GSE83717, GSE99621) and mouse datasets (GSE18800, GSE40151, GSE34814). The algorithm was provided with normalized, z-transformed bulk expression values in order to create a signed network based on biweight midcorrelation. To reduce noise, network’s sign was calculated by adjacency transformation. Network modules were defined by a (1-TOM) measured distance (TOM: Topological Overlap Matrix) and closely related modules – defined by 0.25 and 0.40 units of distance according to hierarchical clustering for human and mouse samples, respectively – were merged. Per module eigengenes (ME) was measured according to default parameters.

Pearson correlation was used to identify any significant ($|\rho| > 0.6$; p -value < 0.05) module-trait-of-interest relationships and intra-modular analysis was utilized to find phenotype drivers. The latter were characterized by both their module membership (MM) and gene expression significance (GS). MM is defined as the Pearson correlation of a gene expression values with the respective ME, while GS as the Pearson correlation of gene expression values with a phenotypic trait vector. Statistical significance of both MM and GS were calculated using *corPvalueStudent* function of the WGCNA R package. Networks that can be plotted in the Fibromine app contain nodes with pre-set MM and GS values above the respective 60th percentile, a threshold that can be interactively modified according to user’s needs. The edges of each network represent correlations of TOM measurements of the 3rd quartile. No genes with zero degree of connectivity are used during minimum spanning tree calculation, while network layout is automatically chosen.

2.10 Pathway analysis

For the needs of PA, two R packages were exploited: *enrichR* (Chen, Tan et al. 2013) and *clusterProfiler* (Yu, Wang et al. 2012). In all cases, significant results were defined by an adjusted p -value < 0.05 . The former tool is a computational interface to Enricher, a heavily-used web-based enrichment analysis tool that hosts multiple reference databases. It is the workhorse of all real-time PA options of Fibromine, thanks to its web-site embeddable character (Fanidis, Moulos et al. 2021). Gene Ontology (GO) database of 2018 was used for both human and mouse datasets. KEGG terms as of 2021 and 2019 were exploited for human and mouse samples, respectively. Last, a collection of COVID-19 related gene sets were also exploited.

For all other PA purposes, either over-representation (ORA) or gene set enrichment analysis (GSEA) were performed with the *clusterProfiler* R package. GO (The Gene Ontology Consortium 2019) and KEGG (Kanehisa, Sato et al. 2019) reference databases were used as provided by the *clusterProfiler* package. On the contrary, MSigDB hallmark sets were accessed via *msigdbR* R package (<https://igordot.github.io/msigdbR/>) (Liberzon, Subramanian et al. 2011). For the needs of pre-ranked GSEA, features were sorted by decreasing FC values. Significantly enriched terms (adjusted p -value < 0.05) were further distinguished into induced or suppressed using positive and negative normalized enrichment score (NES) values, respectively.

2.11 Multiplex ELISA data analysis

Luminex data were examined for the purposes of (Magkrioti, Antonopoulou et al. 2022). During pre-processing performed by an independent analyst, abundance values were transformed to FCs defined as the ratio of the response to a specific stimulus divided by the response to plain medium treatment. A signal was defined as active if it was characterized by a $FC \geq 1.5$, a threshold selected via sensitivity analysis.

For both phosphoproteins and secreted factors, stimuli not causing any activation were filtered out in addition to signals not reaching the significant activation threshold under any treatment. For clustering to take place, FCs were transformed to binary values with 1 and 0 representing activation and non-activation cases, respectively. Distances were calculated using the Gower's metric as implemented in the *proxy* R package (<https://cran.r-project.org/web/packages/proxy/index.html>) so as to reflect related pairs of stimuli. The distance matrix was then used to perform divisive clustering as implemented in *cluster* package (<https://cran.r-project.org/web/packages/cluster/index.html>).

2.12 Machine learning

2.12.1 Data pre-processing

ML was used to classify lung samples to either an IPF or a control phenotype based on gene expression data (Fanidis, Pezoulas et al. 2023). Towards that goal, expression data of the consensus DEGs (cDEGs) from seven (GSE10667, GSE24206, GSE48149, GSE47460_GPL6480, GSE53845, GSE83717, GSE99621) IPF_vs_Ctrl lung datasets were accessed via Fibromine. The datasets were selected according to a set of dataset benchmarking characteristics described in (Fanidis, Moulos et al. 2021). cDEGs were defined as those features being significantly (absolute $FC \geq 1.2$; p -value < 0.05) deregulated towards the same direction (up or down) in a minimum of four of the selected datasets. Consensus fold change ($FC_{consensus}$) was used as a single value representation of each feature's deregulation pattern during fibrosis (Fanidis, Moulos et al. 2021). In the case of GSE24206 advanced IPF samples were used to represent the pathological state.

With the aim of preventing overfitting, semantics similarity was used to identify a smaller number of PF-relative features. DOSE R package (Yu, Wang et al. 2015) was utilized to search for the features most related to the *pulmonary fibrosis* term (DOID:3770). The produced per feature similarity score (SS) was then multiplied by $FC_{consensus}$ to create an integrated score $S_i = SS \times FC_{consensus}$ for gene ranking similarly to (Maghsoudloo, Azimzadeh Jamalkandi et al. 2020). Top 200 ranked genes (half up- and half down-regulated) were selected for downstream processes.

Next, normalized expression values retrieved from Fibromine were standardized via vanilla R functions in order to lie within a unified numerical range and missing values were replaced by zero. This pre-processing was applied in all training/testing (the aforementioned seven) and validation (GSE32537, GSE47460_GSE14550) collections. Datasets were then compared and common genes were maintained and intersected with the 200 ones prioritized by semantics similarity. A conclusive list of 172 transcriptomics elements was used for model training. Classification variable levels (IPF or control) were hot-encoded.

2.12.2 Sex-specific expression examination

Samples hierarchical clustering with complete linkage was performed based on pre-processed expression data of the selected 172 features. Sex metadata were available for 62.5% of the

training/testing and 100% of the validation samples. Cases without any sex annotation were not taken into consideration during clustering.

2.12.3 Tuning, training and evaluation

The task of phenotypic classification was assigned to XGBoost models with a binary hinge objective function. Training/testing and evaluation were all performed using the python packages *sklearn* (Pedregosa, Varoquaux et al. 2011) and *xgboost* (Chen and Guestrin 2016). A Monte-Carlo cross validation (MCCV) with a 75:25 train:test split iterated ten times was used for model training and testing. Tuning of hyperparameters was performed per model making use of a stratified 5-fold approach with random shuffling on each training set followed by a grid-search cross-validation. The following hyperparameters were tuned:

- `n_estimators`: the number of used decision trees
- `learning_rate` (eta): a regularization variable step-wise adjusting feature weights during boosting
- `max_depth`: the maximum allowed depth per tree
- `reg_alpha`: the Manhattan distance corresponding to L1 regularization parameter
- `reg_lambda`: the squared Euclidean distance corresponding to L2 regularization variable

For models' performance evaluation six metrics were used: accuracy, precision, sensitivity, specificity, F1-score and Matthew's correlation coefficient (MCC). While F1-score assesses precision-sensitivity tradeoff, MCC encapsulates the confusion matrix of each model in a single numerical value. The median value across all ten models for each evaluation metric was reported. Performance evaluation took place for both the testing and the independent validation dataset.

2.12.4 Model explanation

Shapley additive explanations, Shapley or just SHAP values were calculated per XGBoost model using the *SHAP* python package (<https://github.com/slundberg/shap>). As suggested by (Lundberg and Lee 2017, Lundberg, Erion et al. 2020) the mean absolute SHAP values were calculated as representatives of features' importance during the classification process. In short, SHAP values measure the contribution of each player (feature) to the outcome of a game (phenotype classification) after considering all possible player coalitions (feature combinations). The coalitions are formed one feature at a time and then feature-specific marginal contributions are averaged in a weighted manner. Thus, based on an input group of M genes (g_1, g_2, \dots, g_M) subtracted from a larger set of L features (g_1, g_2, \dots, g_L) with $M \leq L$, the SHAP value of any given $g_j \in G(S_j)$ is given by the following formula:

$$S_j = \sum \frac{|G|!(P - |G| - 1)!}{P!} (f_g(G \cup \{g\}) - (f_g(G)))$$

where $|G|$ is the number of features included in G and $f_g(G)$ the expected value of the function conditioned on P .

2.12.5 Biomarker lists comparison

ML was also used for the comparison of proposed IPF biomarker lists proposed here with other already-published computationally crafted ones. XGBoost algorithm was trained on the samples referred in section 2.12.3. The maintained features were given by the intersection of the already-published marker sets to the 172 cDEGs detected in all Fibromine-retrieved

datasets of the section 2.12.1 Models were tuned, trained, and tested as described above. Correlation of spirometry and normalized gene expression values was performed using Spearman’s correlation test as implemented in *cor.test* vanilla R (v.4.0.4) function. An absolute $\rho \geq 0.5$ and a p-value < 0.05 were set to define significant relationships.

2.13 Ranking aggregation

A couple of the best-performing ranking aggregation methods (Wang, Law et al. 2022), MAIC (<https://github.com/baillielab/maic>) (Li, Clohisey et al. 2020) and BIRRA (https://github.com/baillielab/comparison_of_RA_methods) (Badgeley, Sealfon et al. 2015), were used in (Fanidis, Pezoulas et al. 2023) for the aggregation of SHAP-prioritized features. During aggregation, every model was treated as a different category. Briefly, MAIC (meta-analysis by information content) is a python-implemented package that works on the hypothesis of datasets sharing some of the genes. Based on them it produces a weight for each experiment and a respective score per gene which is used to summarize evidence and produce a final rank (Li, Clohisey et al. 2020). On the other hand, BIRRA (Bayesian Iterative Robust Rank Aggregation), an R-implemented package, iteratively calculates aggregated rankings from dataset-specific Bayes factors and then utilizes the ranking to modify Bayes factors calculation.

Apart from the aforementioned two, an *ad hoc* ranking aggregation method was created extending the popular majority voting technique. During SHAP-weighted majority voting, genes were ranked by decreasing Shapley value weights. The final rank of each gene i was given by:

$$\sum_{m=1}^M \frac{n_m}{r_m}$$

where m is any given model, n_m the number of features having a non-zero mean absolute SHAP value in m and r_m the rank of the gene of interest in the model m post to genomic features ranking in decreasing order of SHAP values.

2.14 Text mining

The same pipeline of scientific literature mining was used in both (Fanidis, Pezoulas et al. 2023) and (Barbayianni, Kanellopoulou et al. 2023) publications. Initially, XML R package (<https://cran.r-project.org/web/packages/XML/index.html>) was used to download PubMed 2022 baseline and form a corpus of abstracts. The corpus was then searched by *rentrez* R package (Winter 2017) for elements mentioning *IPF*[All Fields] OR ("pulmonary fibrosis"[MeSH Terms] OR "pulmonary fibrosis"[All Fields]) OR ("lung diseases, interstitial"[MeSH Terms] OR "interstitial lung diseases"[All Fields] OR "interstitial lung disease"[All Fields]). *pubmed.mineR* (Rani, Shah et al. 2015) was utilized to perform human gene atomization based on HUGO symbols on the above selected texts. In the case of (Barbayianni, Kanellopoulou et al. 2023) recovered genes were mapped to their mouse deregulated homologues via *biomaRt* (Durinck, Spellman et al. 2009).

2.15 Transcription factor analysis

DoRothEA R package (Garcia-Alonso, Holland et al. 2019) was used for transcription factor (TF) analysis in (Barbayianni, Kanellopoulou et al. 2023). Beginning from mouse regulons, focus was given on high quality interactions that included Quant-Seq DEGs as targets (see section 2.5.2). Further filtering maintained those TF-target pairs in which both members were found deregulated via Quant-Seq. Mode of regulation was used to report opposite and same

direction interactions for TFs acting as repressors and activators, respectively. Network visualizations were crafted using *igraph* (Csárdi and Nepusz 2006) and *visNetwork* (<https://github.com/datastorm-open/visNetwork>) R packages.

2.16 Connectivity Map analysis

CMap analysis was used for drug repurposing purposes in (Barbayaanni, Kanellopoulou et al. 2023). Top 300 deregulated features (half over-, half under-expressed) were selected from NGS data (see section 2.5.2) covering the maximum size per query signature permitted by <https://clue.io/query>. The uploaded signature was compared to that of perturbagens included in Expanded CMap LINCS Resource 2020 (L1000 expression data last updated on 11/23/2021). *cmapR* package (<https://github.com/cmap/cmapR>) was used to isolate compound (*trt_cp*) and peptide (*trt_lig*) signatures with significant similarities (FDR-corrected p-value < 0.05), while signed normalized connectivity score (NCS) was utilized to separate signatures of similar from those of opposite phenotypes. The final signatures list was shaped after filtering out perturbagens that: had an unknown mechanism of action, were deduced from a single biological replicate and/or by treatments lasting less than 24 hours.

2.17 Fibromine-hosted proteomics data handling

Literature was scrutinized for publications describing changes in protein abundance between IPF and control individuals. As the identified proteomic datasets were far fewer compared to transcriptomics (see sections 2.2 and 2.4), the process of final selection was looser for the former compared to that of the latter. Thus, differential proteomic results in tabular format were retrieved from respective publications. Proteins quantified in mixtures/same aptamer were filtered out. Differential abundance thresholds of the original publications were maintained *as were* in order to ensure the already small total number of 693 deregulated proteins.

2.18 Fibromine database creation

Fibromine omics data and third-party annotation were organized in an SQLite database. R packages *RSQLite* (<https://cran.r-project.org/web/packages/RSQLite/index.html>) and *DBI* (<https://cran.r-project.org/web/packages/DBI/index.html>) were used for its creation.

Fibromine database was designed to host normalized read counts, differential expression/abundance data, as well as information from many third-party annotation sources necessary for omics data characterization. Omics publications selected for inclusion were manually scrutinized along with the respective repository entries to glean dataset-specific metadata, such as technology used, tissue sampled, samples sex and any clinical data available. Metadata, such as per condition sample size, were modified to accommodate changes introduced during re-analysis including outlier removal.

Molecular level annotation was retrieved from publicly available databases and processed using *ad hoc* created scripts. Gene-level information including biotype, gene name aliases and chromosome position was retrieved from Ensembl using *biomaRt* package (Durinck, Spellman et al. 2009). Homology pairs between *H.sapiens* and *M.musculus* were also utilized to support transcriptome mapping between animal models and actual disease samples. In order to be able to combine miRNA and mRNA transcriptomic datasets hosted in Fibromine RefSeq (O'Leary, Wright et al. 2016) mRNA, ncRNA and peptide IDs were fetched. GO database (The Gene Ontology Consortium 2019) entries' ID, name, namespace and definition were retrieved from the *go-basic.obo* file removing the obsolete terms.

Data retrieved from miRBase (Kozomara, Birgaoanu et al. 2019) and miRDB (Chen and Wang 2020) database were used for the annotation of non-protein coding RNA datasets. In detail, miRBase accessions and identity numbers, as well as full molecular names were maintained for both mRNA and pre-miRNA molecules. Evidence category and location were additionally extracted for mRNAs. Predicted miRNA-mRNA pairs were downloaded from miRDB for the non-coding molecules included in miRBase.

Protein-level annotation was collected via UniProt/Swiss-Prot (The UniProt Consortium 2018) and STRING (Szklarczyk, Gable et al. 2019) databases. UniProt human data were accessed via the *uniprot_sprot_human.dat.gz* file and mouse data were isolated from *uniprot_trembl_rodents.dat.gz*. For each species protein characteristics, such as length, function, subcellular-localization, disease-associations and technical details including accession codes, identity numbers and entry review status were isolated using custom made scripts. Full protein links, actions and info files for *H.sapiens* (NCBI taxonomy ID 9606) and *M.musculus* (NCBI taxonomy ID 10090) were fetched from STRING database and intersected with data from UniProt/Swiss-Prot so as to focus on common protein molecules.

Last, Fibromine database hosts TF annotation for both species. Human TFs as described in (Lambert, Jolma et al. 2018) were retrieved from <http://humantfs.cabr.utoronto.ca/> (v.1.01). Only factors with known motifs were incorporated in the database. Murine TFs were found at <http://bioinfo.life.hust.edu.cn/AnimalTFDB/#/> (Hu, Miao et al. 2019).

2.19 Protein-protein interaction networks

Each of the Fibromine interactive PPI networks revolve around a queried protein and consist of two interaction shells. Protein interactions depicted are sourced from STRING database data stored into Fibromine database. The first shell contains maximum nine proteins (nodes) that are confidently (interaction score > 700) known to interact (edges) with the queried one. The second shell is comprised of the two most trustworthy interactors for each of the first level molecules. No protein can be simultaneously found in both layers. Network visualization is crafted based on a DrL layout weighted by protein interaction scores.

Fibromine networks can be color-annotated according to differential gene expression data that can describe a variety of transcriptomic comparisons. Initially, each protein (node) is matched to its coding element using UniProt/Swiss-Prot data and in cases of a 1:many protein:gene mapping, the gene characterized as primary by UniProt is maintained. Nodes not included in Swiss-Prot are marked as “Unknown” and colored jorjy blue, the default color of the network. Subsequently, DEA data of the requested comparison are retrieved for each of the remaining nodes from all respective datasets. A consensus direction of deregulation is shaped as thus:

- if a gene is found deregulated in either a single or a couple of datasets in a consistent manner its direction of deregulation is kept *as is*
- if a gene is found deregulated in a couple of datasets with an inconsistent direction of deregulation, the larger dataset’s direction of deregulation is adopted
- if a gene is found deregulated in multiple datasets, the most frequent direction of deregulation prevails.

According to the above pipeline, a gene, and thus a protein node, can be characterized as “Non DE”, “Upregulated” and “Downregulated”, categories that are color-coded grey, red and aqua color, respectively. Default DEA thresholds were set to an absolute FC ≥ 1.2 and a p-value

<0.05, but they can be interactively modified through Fibromine. Statistics corrected for multiple testing bias are presented at the explorer's *Data used* tab.

2.20 Accompanying wet laboratory experimentations

No wet laboratory experimentations described in this thesis were performed by the PhD candidate himself. Details about the materials and methods used, please, refer to the respective publications.

3. Results

3.1 A centralized resource for pulmonary fibrosis omics

3.1.1 Datasets collection

To identify all IPF-related transcriptomic datasets we have exploited multiple online resources. First, 'IPF' and 'bleomycin' search terms were used for the collection of human- and animal-related PubMed publications, respectively. Next, the same keywords along with the 'Transcriptomics' filter were applied in omicsdi.org search. To supplement the aforementioned findings we referred to previous listings and reviews, such as (Villaseñor-Altamirano, Moretto et al. 2020) and (Vukmirovic and Kaminski 2018). Last, ReGEO (Chen, Ramírez et al. 2019) offered an alternative search portal for GEO-deposited datasets not coupled to a Digital Object Identifier (DOI) accompanied publication. From the detected microarray and RNA-seq datasets, we maintained only those that provided raw data via open-access repositories and that were comprised by at least three biological replicates per experimental condition. We also filtered out non-commercial arrays as their technical and biological annotation were most often unavailable. Unique experimental setups, such as that of GSE31934, were the only reason for bypassing the previous filtering step. Last, datasets with only pathological or control/healthy samples were also filtered out, as they cannot by default support any differential expression analysis.

Human IPF versus control proteomics datasets were also identified via careful literature search. The only criterion for their inclusion was the provision of differential abundance results for multiple proteins in the form of a table. No other filtering criteria were used due to IPF-related proteomics data scarcity (Norman, Moore et al. 2018, Khan, Dasgupta et al. 2021).

3.1.2 Transcriptomics datasets re-analysis

Following their collection and initial selection, previously published and *ad hoc* developed pipelines were used for transcriptomics data analysis. To avoid introduction of unwanted variation via differences in applied pipelines, we re-analyzed microarray and RNA-seq datasets in the most consistent manner possible (please refer to Materials and Methods for a detailed description). Briefly, we consistently pre-processed same platform array datasets, while closely related analytical methods were applied across platforms whenever possible. As far as RNA-seq data are concerned, we followed the same alignment strategy with only small changes in read mapping parameters following differences in reads strandness and library type. Subsequently, data curation was performed both manually and computationally, with outlier detection and proper per condition samples separation consisting our main goals. Indicatively, we recruited PCA/MDS plots, arrayQualityMetrics results (Kauffmann, Gentleman et al. 2009) and metaseqR2 quality plots (Fanidis and Moulos 2020) to exclude low quality datasets and samples, as well as outliers, from downstream analyses.

Regarding gene-level DEA, the applied strategy was obligatory dictated by the particularities of each technology. For expression arrays we have used the limma moderated t-test statistics method (Ritchie, Phipson et al. 2015) and if necessary, we have sometimes incorporated additional sources of variation in the model to account for potential batch effects. For RNA-seq datasets, we have used PANDORA (Moulos and Hatzis 2015) as implemented in metaseqR2 R package (Fanidis and Moulos 2020). After retrieving the results of nine individual statistical methods we combined their p-values using PANDORA, in order to reduce false positives contamination and reap certain advantages including better representation of transcript types, such as lncRNAs, and better handling of low count reads (Fanidis and Moulos 2020).

In total and post to low quality datasets exclusion, we have selected and re-analyzed 47 human and 13 murine datasets, spanning multiple conditions and sampling sites/tissues (Table S1).

3.1.3 Data organization

Our next objective was the efficient organization and annotation of the re-analyzed omics data. Towards that purpose we created Fibromine, a database designed to host both transcriptomics and proteomics data, samples metadata and more importantly features annotation (Figure 1). Indicatively, two separate database tables were dedicated to the DEA results of protein-coding and non-coding RNAs, hosting fields such as gene names, FC under specific comparisons and the statistical significance of each result. Another table holds protein differential abundances. Last, normalized transcriptomics values were stored in another table for quick future retrieval.

Subsequently, we designed a collection of phrases capable of separating differential expression comparisons spanning distinct tissues, cell lines and phenotypes deposited under the same dataset id. Each comparison is encoded as a three part phrase: *A_vs_B*. The first and the third component describe the juxtaposed conditions holding information such as sample treatment, while *_vs_* element separates the two states. Usage of the capital letter D accompanied by a number signifies the day of sample collection with respect to a treatment's beginning. Likewise, as all Fibromine-hosted proteomic datasets examine differences between the human pathology and control state, we have used the *IPF_vs_Ctrl* phrase for their description.

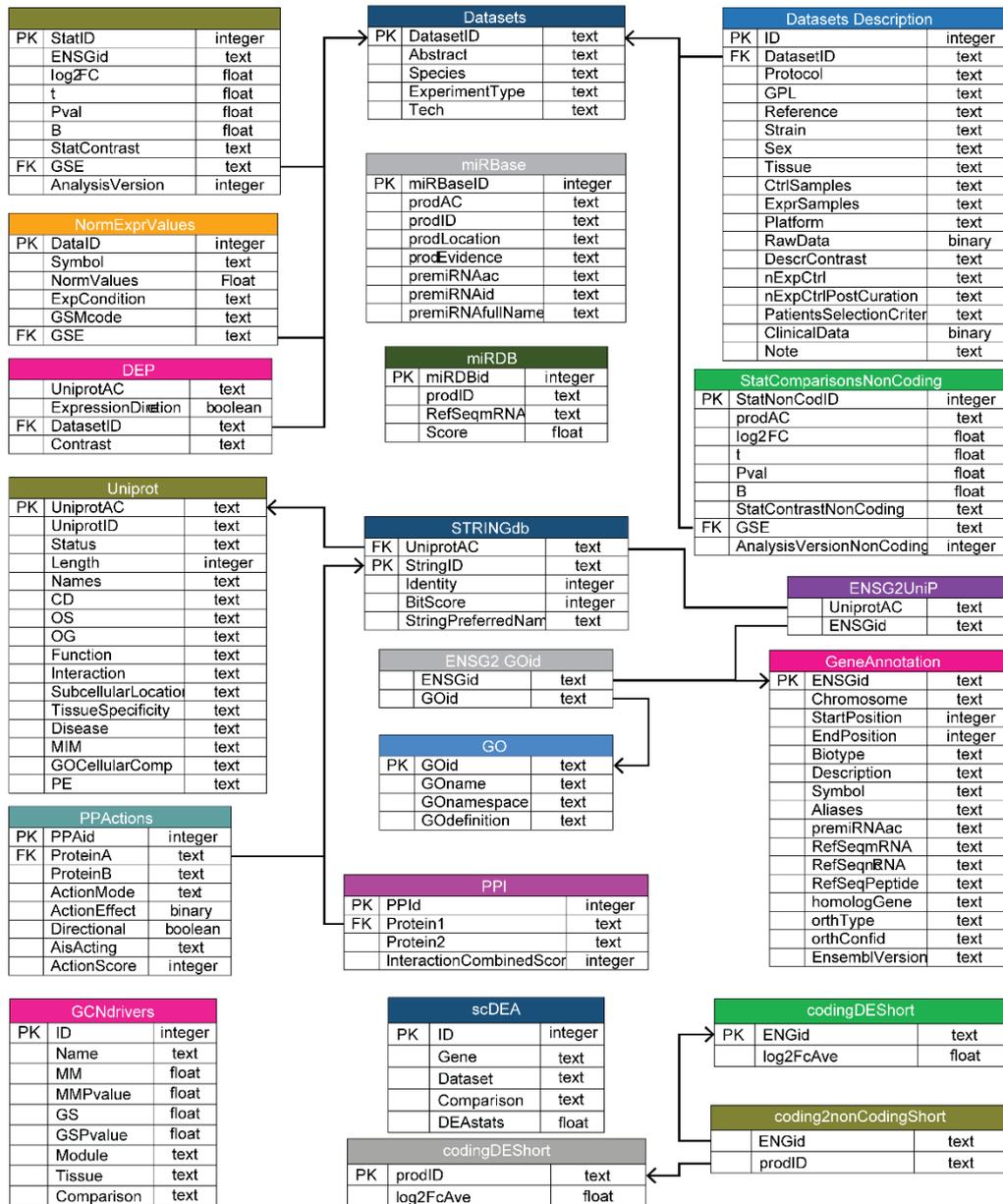


Figure 1 Fibromine database schema. Adapted from (Fanidis, Moulos et al. 2021).

Sample metadata gleaned from both respective publications and supplementary tables accompanying dataset public depositions were also included in two database tables. Information collected during data re-analysis, such as the final dataset size post to outliers/ low quality samples removal and the presence or not of clinical data, were hosted in the same tables. Clinical data *per se* were not incorporated into the database mainly due to their scarcity that rendered them practically useless during any future dataset combination/integration.

To annotate the molecular features of the database, we have collected and re-organized annotation from publicly accessible databases. For transcriptomics data we retrieved gene and transcript-specific content from Ensembl (Yates, Achuthan et al. 2020) and RefSeq (O'Leary, Wright et al. 2016) databases, while miRDB (Chen and Wang 2020) and miRBase (Kozomara, Birgaoanu et al. 2019) were used for small non-coding RNAs description and interconnection with verified or predicted mRNA targets. Protein-related annotation was mined from UniProt (The UniProt Consortium 2018) and STRING (Szklarczyk, Gable et al. 2019) databases. Last,

to functionally characterize the genomic features we have exploited GO database post to obsolete terms removal (The Gene Ontology Consortium 2019).

Collectively, all re-analyzed datasets were organized in a dedicated database along with rich annotation and an *ad hoc* designed comparison vocabulary facilitating data navigation and mining.

3.1.4 Transcriptomics benchmarking

During the aforementioned transcriptomic datasets organization, it became obvious that there was not any metric available for their direct comparison. To accommodate for this lack, we designed and implemented a dataset credit system based on seven criteria, both technical and biological (Figure 2). This system was designed on the assumption that the most accredited datasets would be the most appropriate to combine/integrate, as their data would share multiple commonalities. The system was applied on four bulk transcriptomics DEA comparison groups shaped according to species and target molecules:

1. all coding datasets
2. all non-coding datasets
3. all coding *IPF_vs_Ctrl* lung datasets
4. all coding *BleomycinD14_vs_Ctrl*

For each DEA comparison in any of the above groups we pre-calculated the number of (1) detected and (2) deregulated genes, (3) the proportion of DEGs known to be fibrosis-implicated, (4) the number of features having a small ($1.2 < x < 2$), intermediate ($2 \leq x < 5$) or large absolute FC ($5 \leq x$), (5) the fraction of up to down regulated features, as well as the area under the (6) nominal and (7) multiple test corrected p-value distributions. Human genes known to be implicated in fibrosis were retrieved from (Vukmirovic and Kaminski 2018) along with their direction of deregulation wherever available. The fibrosis-related gene list was enriched with genes sourced from our lab expertise in the field. Mouse orthologues were used for murine datasets evaluation.

Subsequently, the calculated values were used to form per-metric per-group distributions. The latter were evaluated and a dataset was accredited when:

- the number of known fibrotic genes found to be DE in the same direction of deregulation with the literature was larger than the median of the distribution
- the number of detected genes was between the 25th and the 90th percentile of the distribution
- the number of DEGs lied in the interquartile range (IQR) of the distribution
- the ratio of up/down DEGs lied in the IQR of the distribution
- the area under the p-value belonged to the IQR of the distribution
- the area under the corrected p-value belonged to the IQR of the distribution
- the number of DEGs in at least 2 of the 3 FC groups (low, intermediate, high) lied between the IQR of the respective distribution.

As a result, each dataset received minimum 0 and maximum 7 stars. The most stars assigned to a dataset the more ‘trustworthy’ it was considered for downstream combinatorial processes. Data from each benchmarking group were incorporated in the database and can be found in Fibromine’s ‘Dataset Benchmarking’ tab.

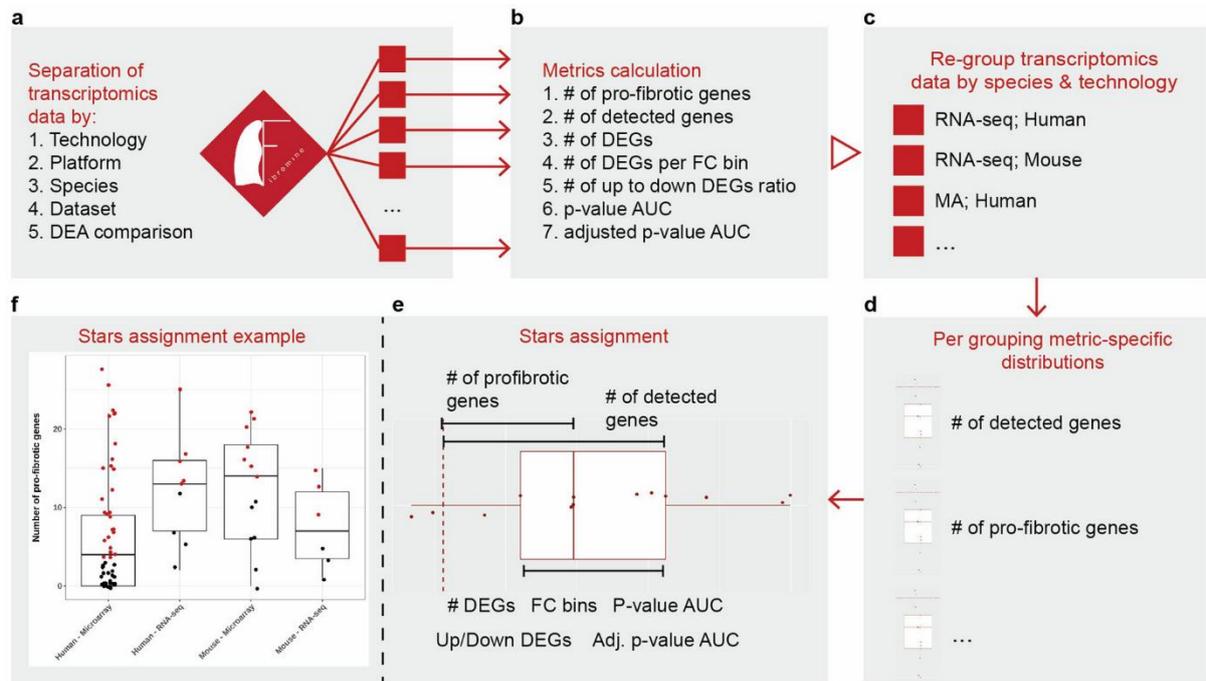


Figure 2 Transcriptomics datasets benchmarking process. Dataset groups are evaluated for seven different metrics, which are then organized in metric-specific distributions. Pre-defined distribution intervals are used to appoint per-dataset stars. The latter are in turn exploited to discern datasets for downstream combinatorial processes. Adapted from (Fanidis, Moulos et al. 2021).

3.1.5 Interactive database access/use

Post to database creation, we set to create an online application offering both data mining and combination services. Named after the database at its core, Fibromine (<https://fibromine.com/Fibromine>) is an R Shiny application designed around three major ('Dataset', 'Gene' and 'Protein' explorer) and several satellite explorers, such as 'Gene co-expression' and 'Datasets benchmarking' (Figure 3). Except for *in situ* data manipulation Fibromine enables download of almost all hosted data supporting downstream analysis.

3.1.5.1 Dataset explorer

'Dataset explorer' is the first main Fibromine interface. It was designed to provide access to differential gene expression and protein abundance data. Selection of any single dataset enables exploration of all its DEA features. On the other hand, multiple transcriptomic datasets selection triggers their combination via identification of consensus differentially expressed genes (cDEGs) (Figure 4). The latter are reported along with their consensus fold change ($FC_{\text{consensus}}$) and consensus direction of deregulation. In species-specific cases, cDEGs are defined by a consistent and same-direction deregulation (up/down) in at least half of the chosen datasets. In parallel, they should lack any significant deregulation towards the opposite direction in the remaining selected cases. During between-species dataset combinations, cDEGs are those features with an 1:1 human:mouse homology and significant consensus expression change in at least half of each species' selected datasets. Post to cDEGs identification, the most frequent direction of deregulation is selected as the consensus, while $FC_{\text{consensus}}$ is set to the mean of the FC values from the datasets matching the consensus direction. User-tunable differential expression thresholds are by default $|FC| \geq 1.2$ and $p\text{-value} < 0.05$. Proteomic dataset combination is achieved via consensus differentially abundant proteins (cDAP) identification (Figure 4). cDAPs are discerned as in transcriptomic datasets

with the exception of original differential abundance thresholds being adopted per dataset. Last, in order to interconnect the two molecular levels, cDAPs are presented along with their respective cDEGs in case of experimental conditions match between selected transcriptomics and the hosted proteomic datasets.

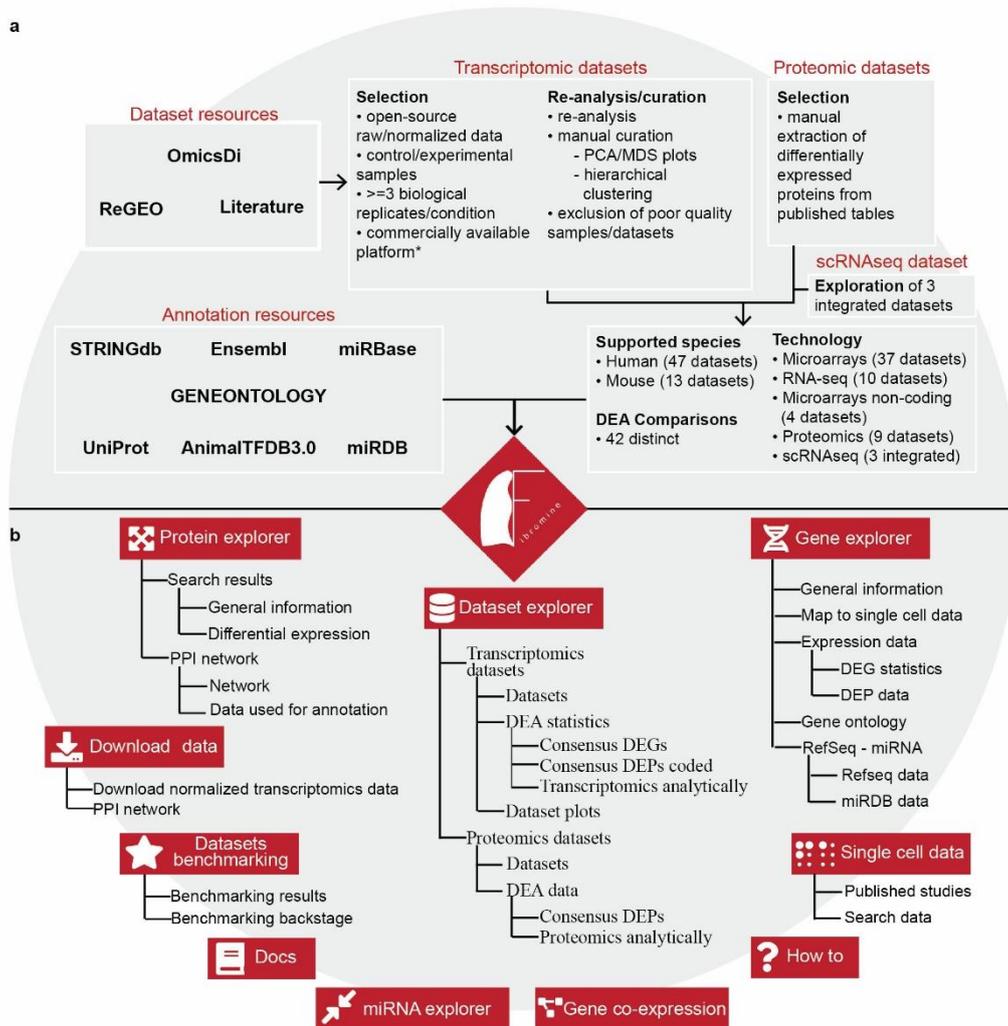


Figure 3. Fibromine schema. A. Fibromine database comprises of multiple omics datasets from various platforms and technologies. Data from open-access databases were used to annotate hosted expression values. B. The homonym application can be used for data mining and integration via multiple explorer tabs. Real time created and pre-formatted data can be freely downloaded. ‘How to’ and ‘Doc’ tabs provide an analytic walk description of the application and its functions. Adapted from (Fanidis, Moulos et al. 2021).

Alongside datasets combination, ‘Dataset explorer’ provides the functionality of over-representation analysis of the projected transcriptomics data (combined or not) via the ‘Pathway analyses’ tool. The latter exploits five reference databases including one of COVID-19-related gene sets. A per-dataset interactive volcano plot and top deregulated genes heatmap can also be visualized on demand.

3.1.5.2 Gene explorer – miRNA explorer – Single cell data

Most of the times, researchers are interested in expression patterns of specific genes rather than whole dataset profiles. To meet such a need Fibromine offers ‘Gene explorer’, its’ second main explorer (Figure 3). Requiring as input a list of HGNC names or Ensembl identification codes,

the explorer displays a series of DE statistics (FC, nominal and multiple testing corrected p-values) for all DE entries across human and/or mouse datasets. Genomic coordinates, RefSeq transcript IDs and related GO functional terms lie among the accompanying annotation. Potential and confirmed miRNA interactions with the queried gene targets are also listed in an attempt to supplement and enrich previous reports (McDonough, Kaminski et al. 2019). Similarly to the ‘Dataset explorer’, ‘Gene explorer’ reports differences in the abundance of protein products coded by the queried genes. Last, in consequence of single cell sequencing emergence, the explorer automatically maps the queried DE features to publicly available scRNA-seq datasets (Reyfman, Walter et al. 2018, Xie, Wang et al. 2018, Joshi, Watanabe et al. 2019). DEA results at the single cell level from another dataset (Mayr, Simon et al. 2021) can be found at the ‘Single cell data’ Fibromine tab, supplementing bulk sequencing data with expression patterns of greater resolution.

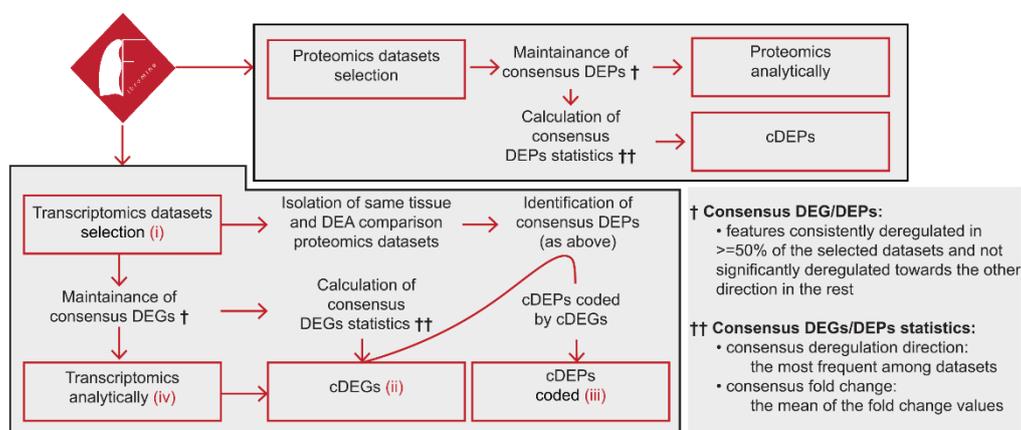


Figure 4 Omics datasets combination pipeline. ‘Dataset explorer’ summarizes DEGs/DAPs across selected datasets through the identification of consensus DEGs/DAPs. Adapted from (Fanidis, Moulos et al. 2021).

Non-protein coding transcripts consist a special category of molecules under both steady state and pathological conditions (Loganathan and Doss C 2023). As in detail recorded in the introduction, miRNAs have many times sparked the interest of the scientific community regarding their role in the fibrotic lung. Fibromine itself hosts several miRNA-oriented expression datasets. Although miRNA expression patterns can be individually assessed via ‘Gene explorer’, it cannot simultaneously provide DE statistics for both members of an miRNA-mRNA interaction. To fill that gap we constructed the ‘miRNA explorer’ hosting data from IPF_vs_Ctrl coding and no-coding datasets (Figure 3). Initially we identified all non-coding cDEGs using default Fibromine parameters. Then, we collected the miRDB-defined targets that have been consensus deregulated towards the opposite direction of deregulation in the respective coding datasets. By combining the above two lists, ‘miRNA explorer’ effortlessly provides multiple miRNA-mRNA pairs that could potentially affect or be affected by PF.

3.1.5.3 Protein explorer

‘Protein explorer’ consists the third core explorer of Fibromine. The explorer accepts as input the name of any single protein coding gene and returns data about its differential abundance during IPF along with UniProt-sourced description metadata. More importantly, the explorer is capable of creating on-demand phenotype-specific PPI networks (Figure 5). These networks

are built on STRING pre-mined protein (nodes) interactions (edges) revolving around the queried protein. They can be annotated with data shared by datasets examining any specific combination of tissue and DE comparison hosted in Fibromine (Materials & Methods). Briefly, following the selection of the above two parameters, the application collects consensus gene expression data from the respective transcriptomic datasets and colors each network node as Non-DE, Up-regulated or Down-regulated. The DEA thresholds used during cDEG identification can be changed manually by the user. Networks are species-specific and follow the queried protein. Thanks to this visual feature, transcriptomics data can be projected to the protein interaction space, offering a potentially significant glimpse into DE impact in the formation of protein complexes and/or in the operation of signaling pathways.

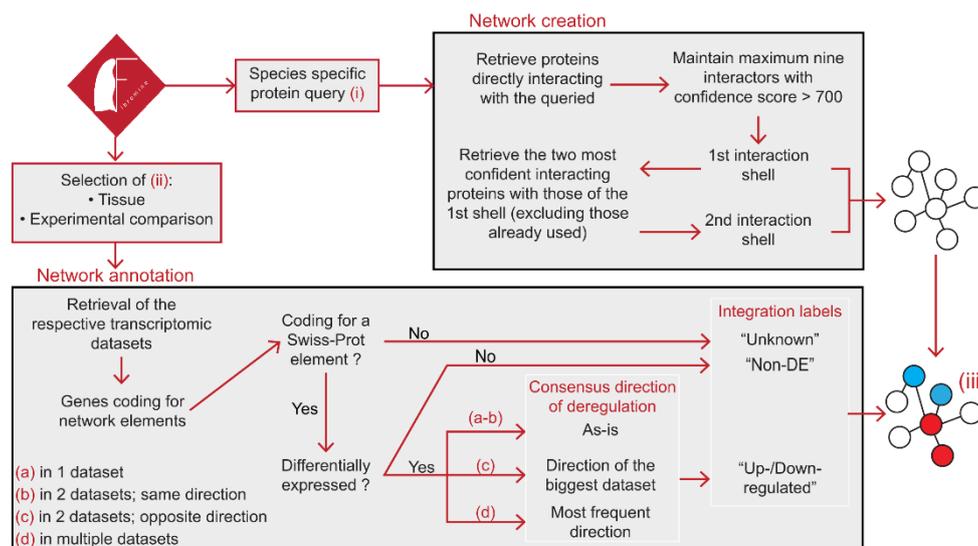


Figure 5 Disease-specific PPI networks creation pipeline. Protein explorer uses transcriptomics data to project on PPI networks the consensus expression profile of a certain tissue/cell type under specific parameters that affect DE. Adapted from (Fanidis, Moulos et al. 2021).

3.1.5.4 Gene co-expression and other smaller explorers

In any given condition, certain gene groups are characterized by the same expression patterns, potentially due to common underlying regulatory mechanisms. These co-fluctuations can be described by gene co-expression networks (GCN) which can, once created, be exploited for various applications, such as functional characterization of molecular features and prioritization of disease targets. Motivated by previous IPF-related publications that have successfully used gene co-expression analysis (McDonough, Kaminski et al. 2019), we implemented ‘Gene co-expression’ Fibromine tab. The latter presents the results of two WGCNAs focusing on human IPF_vs_Ctrl and mouse BleomycinD14_vs_Ctrl lung datasets apiece. Selected human datasets were decorated with more than four stars when benchmarked in the respective lung group of coding expression sets. The three most accredited mouse datasets among same condition sets were used for the murine WGCNA. Briefly, for each analysis, we shaped gene co-expression modules on normalized and standardized gene expression values (Materials and Methods). Subsequently, and to prioritize potential pathology drivers, we filtered in those module members characterized by statistically significant MM and GS metrics and organized them in Fibromine database. With the selection of any of these features from a drop-down list, WGCNA modules are accessible via Fibromine application as interactive GCNs. Network nodes represent statistically significant module members with an

MM and GS value above the 60th percentile of the respective distribution. The latter default threshold can be dynamically tuned for each metric. Network edges reflect strong TOM values (above the 75th percentile of the respective distribution) between all network nodes with the occasional exception of the queried feature.

Last, Fibromine consists of several smaller tabs with further data mining and downloading functionalities, as well as documentation on the use of the application. The ‘Single cell data’ explorer hosts DEA data from (Mayr, Simon et al. 2021) as previously mentioned, along with a list of multiple scRNA-seq IPF datasets. ‘Datasets benchmarking’ tab thoroughly describes the transcriptomics datasets credit system by cataloging all stars and metric distributions. In addition it records all benchmarking results (see section 3.1.4). ‘How to’ and ‘Docs’ tabs provide guidelines on the use of Fibromine and details regarding its features, respectively. Through ‘Download data’ the user can retrieve normalized gene expression data for local downstream analysis.

To conclude with, we have created a collection of carefully selected, consistently re-analyzed and manually curated omics datasets examining IPF. These data have been organized in a dedicated database along with rich third-party annotation and meta-data. As an IPF-oriented reference point the database has been made accessible through the creation of Fibromine, an online, free and easy-to-use application. The latter offers multiple data mining and combination features that require zero computational experience for their successful operation. To the best of our knowledge, our application also hosts the first PF-specific PPI network annotation tool. Ultimately, Fibromine is expected to accelerate IPF research through fast *in silico* hypothesis validation and new hypothesis formation.

3.2 Machine learning-based disease marker prioritization

Fibromine is a great resource for interactive exploration of good quality, manually curated lung fibrosis high-throughput data. Significantly, it enables features prioritization by combining genes or proteins individually ‘endorsed’ by multiple independent datasets (cDEGs/cDEPs). Although any user can exploit Fibromine to export cDEG lists with pathology-related molecular elements (Fanidis, Moulos et al. 2021), selecting a targets shortlist for downstream experimentations is not always straight-forward. To address this limitation and showcase Fibromine’s significance for IPF research, we trained a ML algorithm to classify samples as control or IPF solely based on gene expression data. Focus was given on the IPF_vs_Ctrl comparison both due to its high frequency among Fibromine and literature DE comparisons, as well as due to its direct interest for IPF research. Interpretation of the ML decisions was subsequently used for target selection.

3.2.1 Successful XGBoost separation of sample phenotypes

To begin with, we set to select a ML training/test set. For this purpose we distinguished seven IPF_vs_Ctrl lung datasets accredited with at least 4 stars during Fibromine dataset evaluation (GSE10667, GSE24206, GSE48149, GSE47460_GPL6480, GSE53845, GSE83717, GSE99621). Fibromine’s ‘Dataset explorer’ was utilized for datasets combination resulting in the identification of 2182 cDEGs. Compared to the total of 184 (124 IPF and 60 control) samples, the dataset was characterized by a disproportionate features-to-examples (genes-to-samples) ratio. To deal with it we exploited a previously published semantics similarity prioritization pipeline (Maghsoudloo, Azimzadeh Jamalkandi et al. 2020) after a series of context-required modifications (Materials and Methods). In a nutshell, we designed and

calculated an integrated score (S_i) that combined both consensus gene expression data and cDEGs semantics similarity to pulmonary fibrosis related terms. S_i was then used to order features, with up-regulated PF-related genes ending at the top and down-regulated PF-related genes at the bottom of the list. Top 100 and bottom 100 features were selected for ML training and testing (Figure 6).

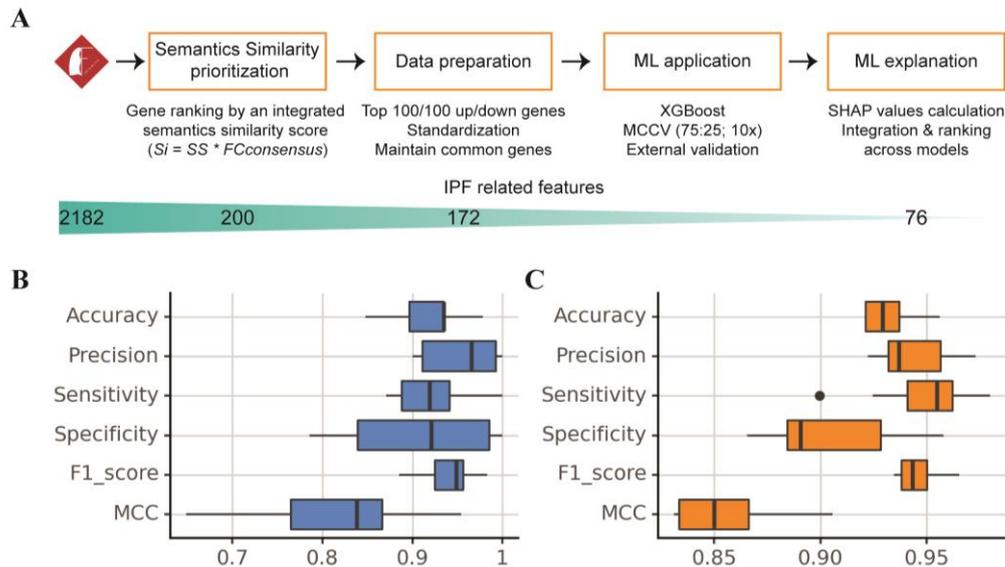


Figure 6 Machine learning methods are capable of separating IPF-originating from control samples using gene expression data. Adopted from (Fanidis, Pezoulas et al. 2023).

Post to samples and features selection, we pre-processed data for proper use with ML methods (Figure 6). Initially, Fibromine-hosted gene expression values from all seven datasets were downloaded and standardized, in order to eliminate between dataset differences in numerical range. Next, those features shared by all seven datasets were maintained and intersected with the 200 S_i -prioritized features. The intersection was extended to include genes found in an independent couple of big IPF_vs_Ctrl lung datasets (GSE32537, GSE47460_GSE14550) which was later exploited as an independent ML evaluation set. The latter two datasets included 199 IPF and 119 control samples and were subjected to the same pre-processing steps with the aforementioned seven. As a result, our full training/testing and evaluation sets consisted of 184 (124 IPF - 60 control) and 318 (199 IPF - 119 control) examples, respectively. Both sets beared expression values from 172 cDEGs (Table S2) equally distributed between up and down regulated groups (84 up and 88 down). Last, due to the sex-specific nature of both gene expression (Gershoni and Pietrovski 2017) and IPF pathology (Raghu, Collard et al. 2011), we tested the selected data for sex-driven patterns, with no findings (Figure 7).

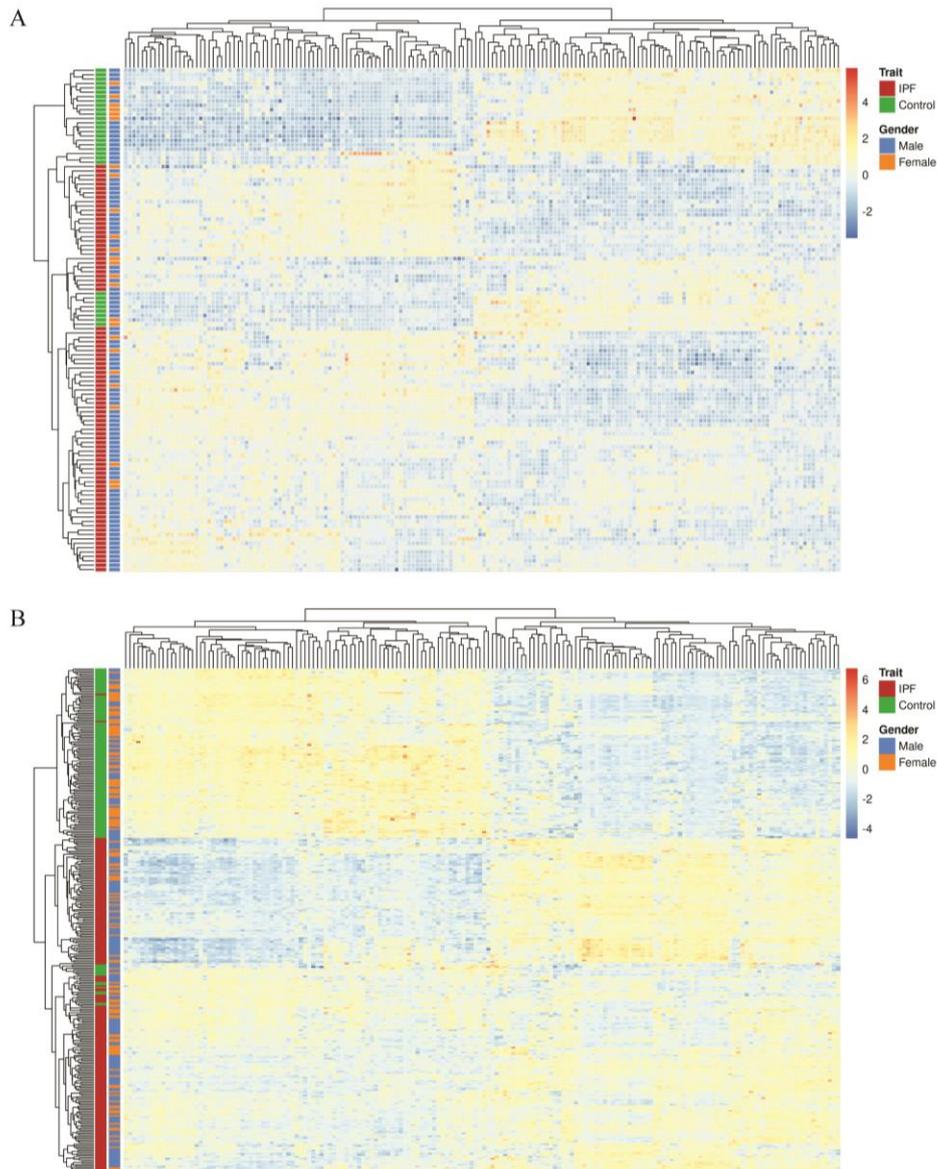


Figure 7 Gene expression data used for ML training and evaluation were not affected by any sex-bias. A. Training/testing dataset. **B.** External evaluation dataset. Adopted from (Fanidis, Pezoulas et al. 2023).

A supervised XGBoost model was chosen to phenotypically classify samples. The learning and evaluation process was based on a MCCV approach with a 75:25 train:test split iterated ten times. Hyperparameter tuning was performed with a stratified 5-fold strategy with random shuffling per training set, accompanied by a grid-search cross-validation (Materials and Methods). Algorithmic performance was summarized by the median across-iterations value for each evaluation metric. As indicated by almost all metrics having a median above 0.9, our algorithm succeeded in samples classification in both test splits (Figure 6B; Table 1) and the independent datasets (Figure 6C; Table 1) More importantly, we observed that ML application on the independent set of data had a smaller between-iterations variation in metric values compared to the initial testing splits (Figure 6C).

Table 1 Machine learning algorithm evaluation. Comparison of our full and lite models with others trained with previously identified target genes. Values presented are the median across training/testing iterations. Colored values represent performance picks. Adopted from (Fanidis, Pezoulas et al. 2023)

Model	Performance metrics											
	Accuracy		Precision		Sensitivity		Specificity		F1-score		MCC	
	Test	Val.	Test	Val.	Test	Val.	Test	Val.	Test	Val.	Test	Val.
Full	0.93	0.93	0.97	0.94	0.92	0.95	0.92	0.89	0.95	0.94	0.84	0.85
Lite	0.93	0.94	0.97	0.95	0.94	0.95	0.91	0.92	0.95	0.96	0.85	0.88
Wan	0.93	0.93	0.95	0.95	0.94	0.97	0.89	0.87	0.95	0.95	0.84	0.86
Li	0.91	0.93	0.95	0.93	0.91	0.96	0.90	0.89	0.94	0.95	0.80	0.86
Zhu	0.92	0.92	0.95	0.92	0.94	0.94	0.89	0.87	0.95	0.93	0.81	0.83
Cui	0.93	0.92	0.97	0.92	0.94	0.96	0.91	0.85	0.96	0.94	0.85	0.83
Xu	0.91	0.86	0.94	0.85	0.95	0.93	0.86	0.74	0.94	0.89	0.80	0.69

Full: the full proposed model; **Lite:** the lite proposed model; **Test:** testing split; **Val:** independent validation set

Conclusively, artificial intelligence efficiently separated samples in an either IPF or control category based on Fibromine-sourced gene expression data.

3.2.2 Identification and functional characterization of ML-utilized features

As ML models were proven capable of separating human phenotypes, we hypothesized that some of the most significant IPF-related transcriptional features were used in the process. To shed light into XGBoost decisions we calculated SHAP values per model and feature (Materials and Methods). As a result, the most informative genes for condition classification were marked by a per-iteration non-zero mean absolute SHAP value (Figure 8A). These 76 features were characterized by a varying frequency of occurrence between the 10 XGBoost models (Figure S1). At the same time it became obvious that feature selection and usage was performed in an iteration-specific manner, with models exploiting a minimum number of 11 and a maximum number of 32 exploited features (Figure S1).

In an attempt to validate the relevance between these 76 genes and PF, we have executed a series of functional and data mining analyses. To begin with, we performed pathway over-representation analysis using GO as a reference database. Some of the most significantly enriched ontologies pertained to ECM and collagen fibrils organization as well as tissue remodeling, suggesting an important role of the prioritized features in lung fibrosis (Figure 8B; Table S3). Similar fibrosis-related sets of genes were revealed when the analysis was repeated with MSigDb hallmark sets as reference (Figure 8C; Table S4). Subsequently, we queried all 76 features in CORUM3.0 (Giurgiu, Reinhard et al. 2019), so as to retrieve protein complexes potentially known for their pro- or anti-fibrotic function. Several of the molecular formations listed have been documented to participate in immune response, integrin signaling, cellular adhesion and migration (Figure 8D; Table S5). Next, we mined a corum of PubMed abstracts flagged by at least one of the ‘pulmonary fibrosis’, ‘IPF’ and ‘interstitial lung diseases’ keywords for gene names. Retrieved genomic elements were then intersected with the 76 ML-used features. Surprisingly, 36 of the latter had already been mentioned in PF-related literature pieces, corroborating selected genes relationship with the fibrotic lung (Figure 8E). Manual scrutiny of the full list further validated the known role of several of its members in lung pathology. Notably, *AGER*, *SPPI1*, *MMP7* and *CRLF1* were included in the (Vukmirovic and Kaminski 2018) list of PF deregulated genes, while 12 of 76 genes, such as *LCN2*, *COL1A1*, *MMP7*, *COL15A1* and *COMP* were chosen by multiple publications as potential drivers of IPF

(Table 2) (Li, Liu et al. 2020, Xu, Mo et al. 2020, Cui, Ji et al. 2021, Wan, Huang et al. 2021, Zhu, Xu et al. 2021).

Table 2. IPF target genes identified by previous publications *in silico* integrating IPF and control gene expression data. Adopted from (Fanidis, Pezoulas et al. 2023).

Prioritized Gene	Xu et al. PMID: 32899090	Zhu et al. PMID: 34522168	Cui et al. PMID: 33672678	Wan et al. PMID: 34476240	Li et al. PMID: 32782692
COL1A1		X	X	X	X
COL15A1					X
COMP		X			
MMP7		X	X	X	
COL17A1					X
SPP1		X	X	X	
ASPN				X	
LCN2		X			
COL3A1		X		X	X
ITGB8			X		
COL1A2	X			X	X
CP					X

3.2.3 Disease targets prioritization

Having validated the fibrosis-relevance of the 76 ML-utilized transcriptomics features, we then set to prioritize them in an informative manner. Towards that goal, we integrated the iteration-specific SHAP-ordered gene lists by means of ranking aggregation. BIRRA (Badgeley, Sealfon et al. 2015) and MAIC (Li, Clohisey et al. 2020) methods were recruited as two of the best available options for handling multiple ranked lists of unknown heterogeneity (Wang, Law et al. 2022). In addition, we applied SHAP-weighted majority voting, an *ad hoc* developed method based on the voting ranking aggregation strategy (Materials and Methods) (Figure 9A).

At first glance, each algorithm seemed to yield a completely different gene ranking, but closest examination revealed that all three aggregations share 65% of their 1st quartile members (Figure 9B). When we examined the three lists in their full length via Kendall ranking correlation, the BIRRA-originating list had the more strong relationship with the other two which in turn shared the less common rankings (Figure 9C). Significantly, 8 of the 13 1st quartile common genes have been reported by the above described text mining process and half of them (*COMP*, *COL1A1*, *COL17A1* and *COL5A2*) were among previously proposed biomarkers (Li, Liu et al. 2020, Cui, Ji et al. 2021, Wan, Huang et al. 2021, Zhu, Xu et al. 2021).

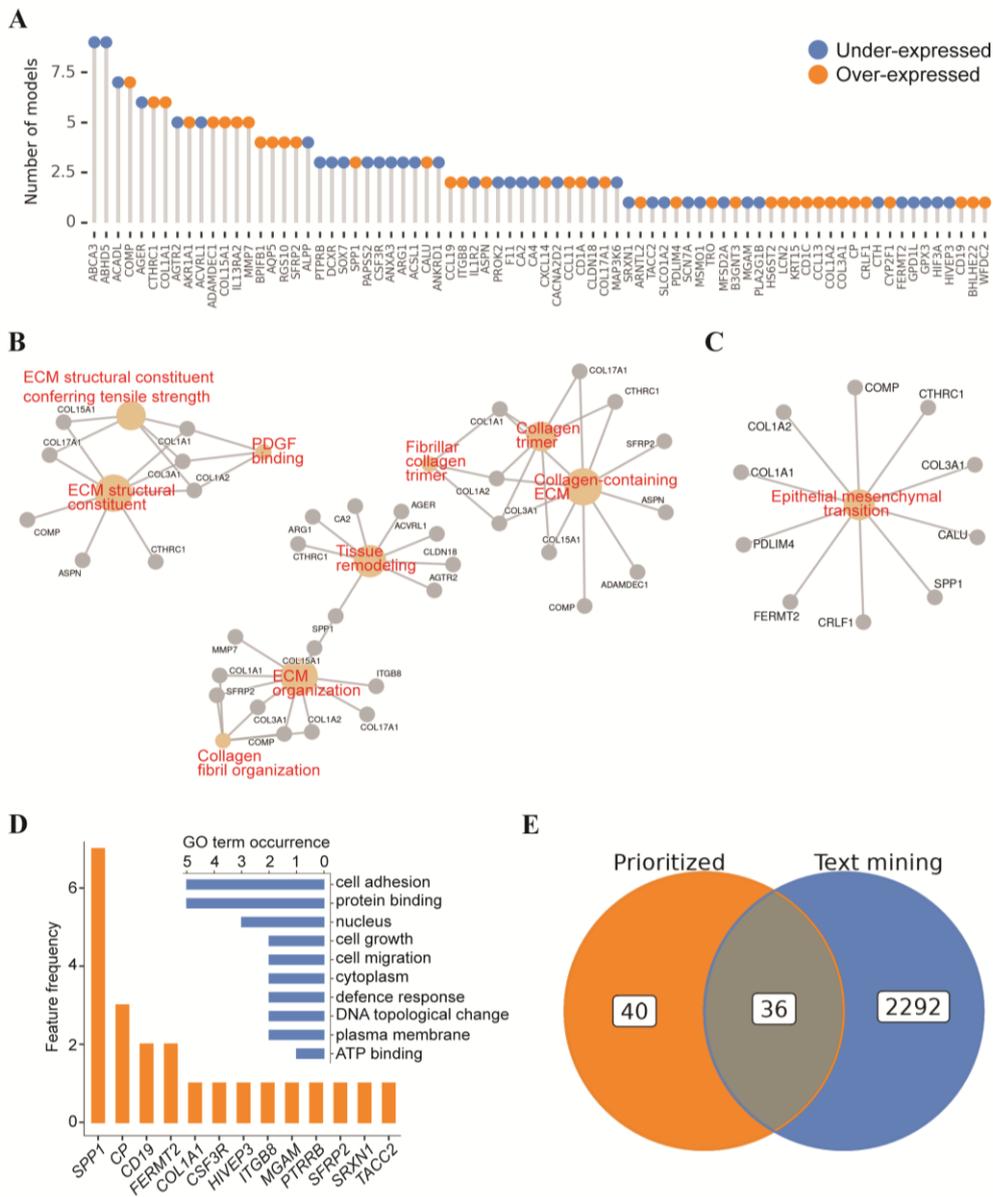


Figure 8 Functional analyses of model-selected features suggest close relationship to pulmonary fibrosis. A. The 76 features exploited by the XGBoost iterations. Direction of deregulation as defined by Fibromine’s ‘Dataset explorer’. **B-C.** GO- and MSigDb-oriented pathway analyses, respectively, highlight involvement in PF-related processes. **D.** Prioritized features participating in CORUM3.0-defined protein complexes along with their functional significance. **E.** PubMed abstract mining reveals the fibrotic character of the identified target genes. Adopted from (Fanidis, Pezoulas et al. 2023).

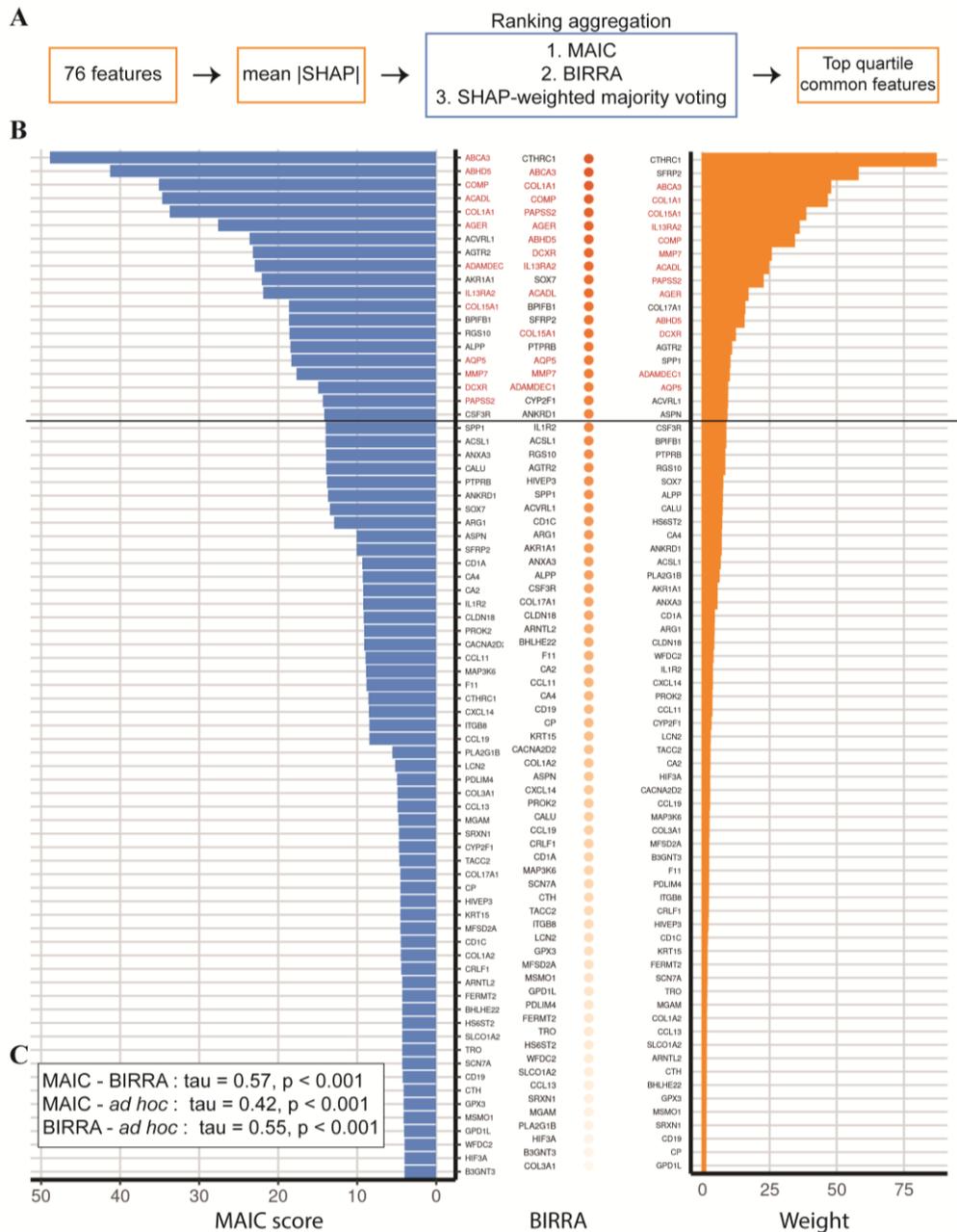


Figure 9 Ranking aggregation of model-prioritized features. A. Ranking aggregation pipeline. **B.** Prioritized features ranking per applied method. Across-methods shared genes of the first quartile are colored in red. **C.** Correlation of the three rankings. That of BIRRA shared the most characteristic with the other two. Adopted from (Fanidis, Pezoulas et al. 2023).

Ultimately, we have applied a ML algorithm on gene expression data and have succeeded not only to distinguish IPF from control samples, but also to use that classification process as a means of disease targets isolation and informed prioritization. We are of the opinion that future research should focus on the less best studied of the 76 prioritized features as latent yet potentially significant IPF players.

3.2.4 Prioritized features list comparison with previous biomarker lists

In the past, several *in silico* pipelines have attempted to identify novel IPF markers using transcriptomics datasets (Li, Liu et al. 2020, Xu, Mo et al. 2020, Cui, Ji et al. 2021, Wan, Huang et al. 2021, Zhu, Xu et al. 2021). Juxtaposition of their findings, revealed minimum common

ground (Figure 10A; Table 2), while comparison with our much longer 76 gene list just validated the underlying heterogeneity. Intrigued by these observations, we decided to examine the above target sets' ability for samples phenotypic classification.

For a proper comparison, XGBoost was once again recruited as a binary classifier using expression values from (Li, Liu et al. 2020, Xu, Mo et al. 2020, Cui, Ji et al. 2021, Wan, Huang et al. 2021, Zhu, Xu et al. 2021) and our prioritized gene list (Figure 8A). In addition, we trained and evaluated a seventh XGBoost algorithm with a more compact version of the full target list. This new gene set, from now on called 'lite', comprised of the top 13 features shared between the three ranking aggregation methods (Figure 9B). Already published target genes not present in the 172 cDEG set were excluded from ML training and testing. As can be seen from Figure 10B and Table 1, our lite model had the best performance next to that of Wan and Cui lists, while all three performed up to 5% better than the rest of the models. On the other hand, Xu-based model had the poorest performance, with an exception in sensitivity. As far as the external dataset evaluation is concerned (Figure 10C; Table 1), the here proposed lite model achieved an overall best classification. In detail, it reached a 92% specificity followed only by the full and Li models, while it also scored the highest accuracy (94%), precision (95%), F1-score (96%) and MCC (88%). Once again, Xu classifier had the least favorable phenotype classification.

Subsequently, we used SHAP values and SHAP-weighted majority voting, in order to explain lite model's *modus operandi* and to prioritize target genes (Figure 10D). Similar to the full model, known fibrotic players, such as *MMP7*, *COL15A1* and *COMP* (Vukmirovic and Kaminski 2018) (Table 2), were placed at the top of the list. *IL13RA2*, a relatively unknown gene in the PF context, not covered by our text mining endeavor was interestingly ranked fourth.

Last, normalized expression values were correlated with spirometry measurements so as to evaluate less studied lite model genes. Both molecular and physiological measurements originated from three cohorts utilized during ML training/testing (GSE47460_GPL6480, GSE47460_GPL14550, GSE32537). Analysis not only revalidated relationships known to literature, but more importantly revealed previously latent ones (Table S6). Illustratively, % predicted Forced Vital Capacity (FVC) was found to be negatively correlated with *IL13RA2* expression in two of the three examined cohorts (Figure 10E), while the opposite pattern was established for *PAPSS2* (Figure 10F).

To wrap up, both our full (76 genes) and lite (13 genes) biomarkers list models were among the top phenotypic classifiers compared with previously published marker sets. Importantly, our lite model had the best performance when evaluated on the external dataset, further validating its usefulness in pathology identification.

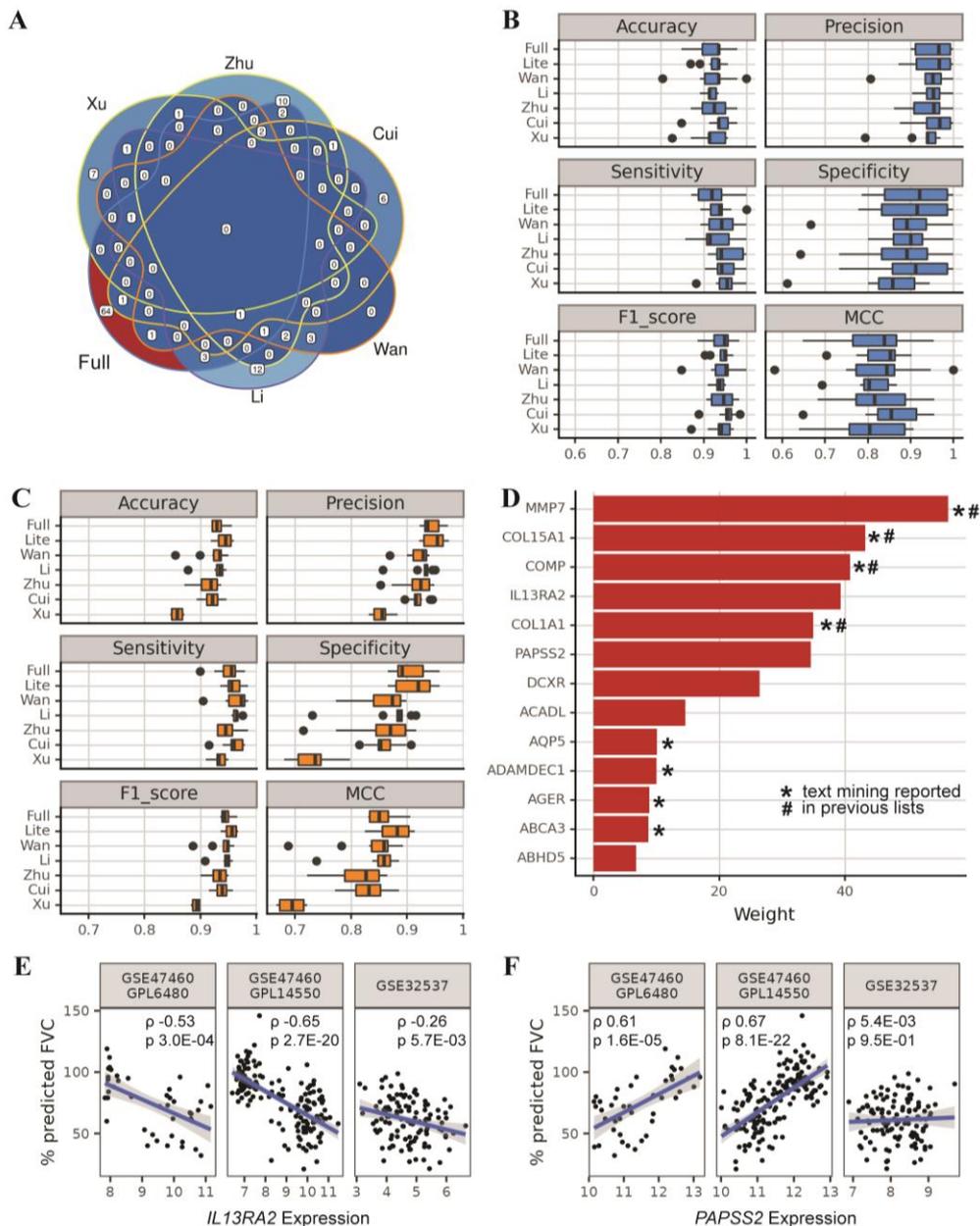


Figure 10 Proposed feature sets performs better than most previously published ones in phenotype classification. **A.** *In silico* proposed disease target collections are highly variable. **B.** Sample classification potential of seven gene sets according XBGoost test splits. **C.** Sample classification potential of seven gene sets according an independent validation dataset. **D.** SHAP-weighted majority voting prioritization of the lite gene list. **E-F.** Correlation of gene expression and spirometry measurements. Adopted from (Fanidis, Pezoulas et al. 2023).

3.3 *ENPP2* expression and methylation in cancer

Apart from its implication in PF (Magkrioti, Galaris et al. 2019), *ENPP2* is a known player in several malignancies (Peyruchaud, Saier et al. 2020). Nevertheless, and despite indications from other diseases (Parris, Kovács et al. 2014, Udomsinprasert, Kitkumthorn et al. 2017), there no systematic investigation of *ENPP2* methylation has been performed.

3.3.1 *ENPP2* methylation in cancerous and healthy tissue

Motivated by the aforementioned data, we set to *in silico* investigate *ENPP2* methylation in tumor and control samples. Initially, in order to reveal a homeostatic baseline of *ENPP2* methylation across the human body, we re-analyzed GSE50192, a dataset spanning 17 healthy tissues (Lokk, Modhukur et al. 2014). As can be seen from (Figure 11), all gene body CGs had elevated levels of methylation when compared to those located at the transcription start site (TSS) and the 1st exon of the gene.

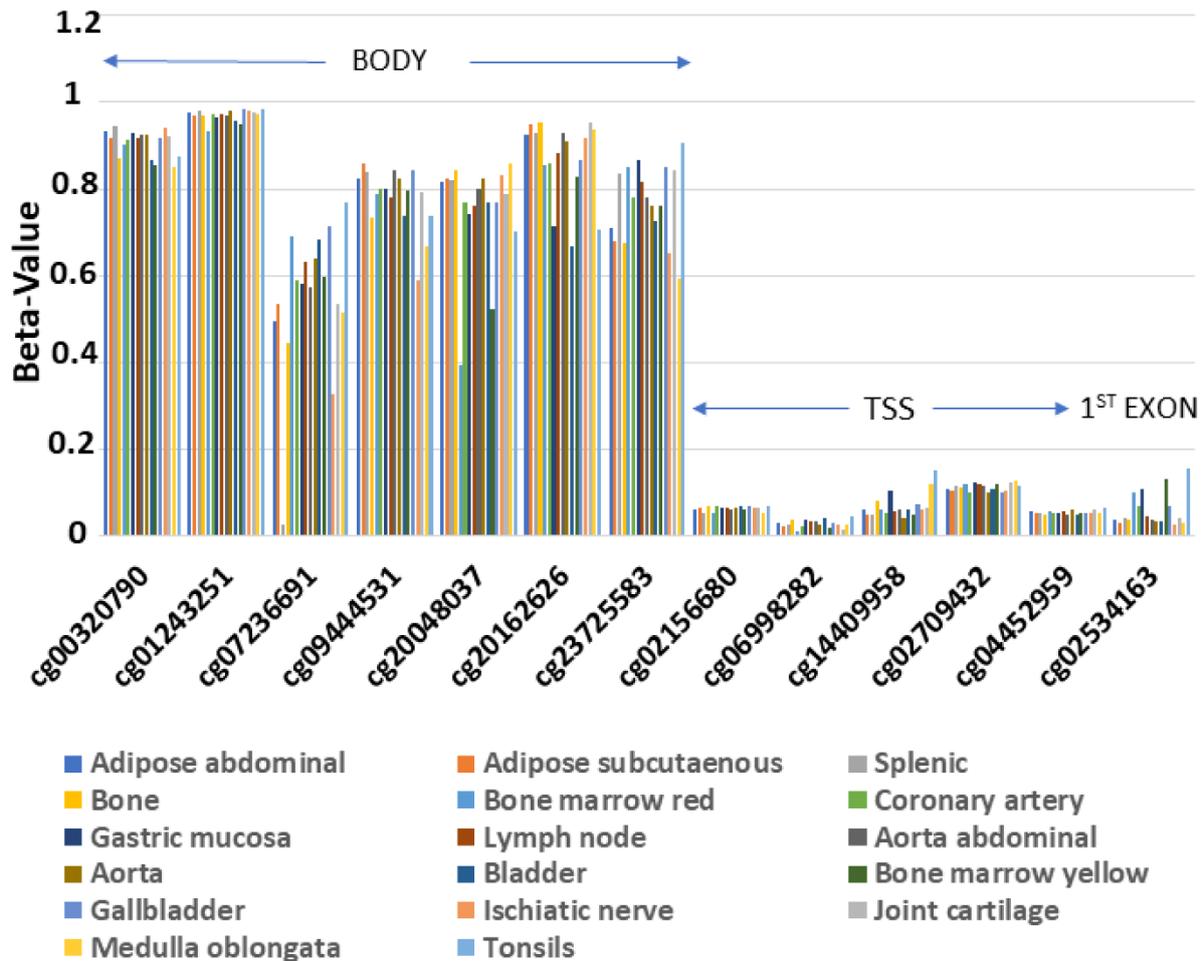


Figure 11 Pattern of *ENPP2* methylation across human tissues. TSS: Transcription start site

Re-analysis of 10 publicly available datasets (GSE113017, GSE113019, GSE120878, GSE27130, GSE63704, GSE76938, GSE98534, GSE46306, GSE134772, GSE97686) investigating DNA methylation in some of the most frequently arising and lethal cancer types (Sung, Ferlay et al. 2021), revealed differentially methylated CGs (DMCs) in all but precancerous interepithelial cervical neoplasia (CIN), cervical cancer (CC) and gastric cancer (GC) cases (Table 3). Several DMCs were recurrently identified. Importantly, all promoter-associated and 1st exon CGs had increased methylation during cancer. The opposite pattern was witnessed in almost all gene body CGs.

To extend our research, we reanalyzed GSE71627, a dataset that contains hepatocellular and prostate cancer cell lines of differential aggressiveness (Cheishvili, Stefanska et al. 2015). 12 DMCs were identified (Table 4) and half of them were shared by both tumor types. With the only exception of two regions, CGs across the whole gene had a heavier decoration in the more

Table 3. ENPP2 differential methylation in various cancer types compared to steady state.
Adopted from (Panagopoulou, Fanidis et al. 2021).

Study ID	Compared Groups	CG ID	M β Value 1 *	M β Value 2 *	$\Delta\beta$ Value #	Regulation	Gene Region	FDR
HCC								
GSE113017	Adjacent liver vs. HCC tumor	cg00320790	0.892	0.761	-0.130	Down	Body	2.5×10^{-2}
		cg07236691	0.832	0.659	-0.173	Down	Body	1.9×10^{-3}
		cg09444531	0.795	0.619	-0.176	Down	Body	8.6×10^{-4}
		cg20162626	0.809	0.642	-0.167	Down	Body	3.0×10^{-2}
		cg02709432	0.123	0.263	0.141	Up	TSS200	1.9×10^{-2}
		cg04452959	0.058	0.189	0.131	Up	TSS200	2.1×10^{-2}
		cg02156680	0.067	0.202	0.135	Up	TSS1500	4.0×10^{-3}
		cg06998282	0.086	0.286	0.200	Up	TSS1500	4.0×10^{-3}
GSE113019	Adjacent liver vs. primary HCC tumor	cg00320790	0.835	0.623	-0.212	Down	Body	3.0×10^{-2}
		cg07236691	0.776	0.563	-0.213	Down	Body	4.5×10^{-2}
		cg09444531	0.721	0.484	-0.237	Down	Body	2.4×10^{-3}
		cg20048037	0.696	0.474	-0.222	Down	Body	2.1×10^{-2}
		cg20162626	0.695	0.447	-0.248	Down	Body	7.0×10^{-3}
		cg23725583	0.718	0.557	-0.161	Down	Body	6.2×10^{-2}
		cg02156680	0.064	0.188	0.124	Up	TSS1500	1.8×10^{-2}
		cg06998282	0.105	0.284	0.179	Up	TSS1500	3.0×10^{-2}
	Adjacent liver vs. (primary & recurrent) HCC tumor	cg02709432	0.127	0.256	0.129	Up	TSS200	2.0×10^{-2}
		cg04452959	0.042	0.149	0.106	Up	TSS200	9.3×10^{-3}
		cg00320790	0.835	0.634	-0.201	Down	Body	2.3×10^{-3}
		cg07236691	0.776	0.568	-0.207	Down	Body	2.3×10^{-3}
		cg20048037	0.696	0.450	-0.246	Down	Body	1.8×10^{-3}
		cg23725583	0.718	0.581	-0.137	Down	Body	4.6×10^{-2}
		cg02156680	0.064	0.198	0.134	Up	TSS1500	4.0×10^{-3}
		cg06998282	0.105	0.279	0.174	Up	TSS1500	1.3×10^{-2}
		cg14409958	0.263	0.374	0.111	Up	TSS1500	2.2×10^{-2}
		cg02709432	0.127	0.253	0.126	Up	TSS200	1.2×10^{-2}
		cg04452959	0.042	0.141	0.098	Up	TSS200	5.3×10^{-3}
		cg02534163	0.143	0.274	0.131	Up	1st Exon	2.3×10^{-3}
Melanoma								
GSE120878	Nevus vs. primary melanoma tissues	cg23725583	0.481	0.583	0.102	Up	Body	4.7×10^{-6}
		cg00320790	0.859	0.834	-0.025	Down	Body	2.3×10^{-3}
		cg09444531	0.736	0.692	-0.043	Down	Body	1.1×10^{-2}
		cg20048037	0.801	0.740	-0.061	Down	Body	1.4×10^{-6}
		cg20162626	0.625	0.575	-0.050	Down	Body	2.3×10^{-3}
		cg26078665	0.647	0.616	-0.031	Down	Body	8.0×10^{-3}
		cg04452959	0.105	0.142	0.037	Up	TSS200	5.6×10^{-4}
		cg02534163	0.155	0.242	0.087	Up	1st Exon	1.4×10^{-6}
CRC								
GSE27130	Adjacent colon vs. CRC	cg14409958	0.201	0.210	0.009	Up	TSS1500	5.6×10^{-3}

Table 3 cont.

Study ID	Compared Groups	CG ID	M β Value 1 *	M β Value 2 *	$\Delta\beta$ Value #	Regulation	Gene Region	FDR
LC								
GSE63704	Normal lung vs. LC	cg09444531	0.832	0.774	-0.058	Down	Body	3.0×10^{-4}
		cg20048037	0.832	0.754	-0.078	Down	Body	5.2×10^{-5}
		cg20162626	0.863	0.802	-0.061	Down	Body	6.5×10^{-5}
		cg02709432	0.234	0.265	0.030	Up	TSS200	1.7×10^{-2}
		cg06998282	0.198	0.268	0.070	Up	TSS1500	2.6×10^{-3}
		cg14409958	0.264	0.366	0.103	Up	TSS1500	1.4×10^{-4}
	cg02534163	0.133	0.183	0.050	Up	1st Exon	1.5×10^{-2}	
	IPF vs. LC	cg00320790	0.940	0.904	-0.036	Down	Body	5.6×10^{-4}
		cg20048037	0.785	0.754	-0.031	Down	Body	5.2×10^{-1}
		cg20162626	0.844	0.802	-0.042	Down	Body	2.1×10^{-2}
		cg06998282	0.196	0.268	0.072	Up	TSS1500	4.7×10^{-3}
		cg14409958	0.269	0.366	0.097	Up	TSS1500	8.4×10^{-4}
	COPD vs. LC	cg20048037	0.810	0.754	-0.056	Down	Body	8.4×10^{-3}
		cg20162626	0.850	0.802	-0.048	Down	Body	4.7×10^{-3}
		cg02709432	0.229	0.265	0.035	Up	TSS200	8.4×10^{-3}
		cg02156680	0.183	0.204	0.022	Up	TSS1500	3.4×10^{-2}
		cg06998282	0.184	0.268	0.084	Up	TSS1500	8.4×10^{-4}
		cg14409958	0.253	0.366	0.114	Up	TSS1500	8.4×10^{-4}
cg02534163		0.133	0.183	0.050	Up	1st Exon	2.8×10^{-2}	
PC								
GSE76938	benign prostate vs. PC	cg07236691	0.471	0.670	0.199	Up	Body	1.4×10^{-6}
		cg09444531	0.654	0.775	0.122	Up	Body	1.4×10^{-6}
		cg23725583	0.845	0.914	0.069	Up	Body	1.0×10^{-3}
		cg26078665	0.687	0.709	0.022	Up	Body	7.6×10^{-3}
		cg20162626	0.787	0.687	-0.100	Down	Body	1.4×10^{-6}
		cg02709432	0.093	0.417	0.324	Up	TSS200	1.4×10^{-6}
		cg04452959	0.034	0.289	0.255	Up	TSS200	1.4×10^{-6}
		cg06998282	0.108	0.457	0.349	Up	TSS1500	1.4×10^{-6}
		cg14409958	0.126	0.399	0.273	Up	TSS1500	1.4×10^{-6}
		cg02156680	0.070	0.351	0.281	Up	TSS1500	1.4×10^{-6}
		cg02534163	0.071	0.340	0.268	Up	1st Exon	1.4×10^{-6}

* Mean β (M β) value 1 represents methylation in control and Mean β (M β) value 2 methylation in cancer; # $\Delta\beta$ value: (Mean β value 2-Mean β value 1). Abbreviations: DMC: Differentially Methylated CpG; FDR: False Discovery Rate; HCC: Hepatocellular Carcinoma; TSS: Transcription Start Site; PA: Promoter Associated; CRC: Colorectal Cancer; IPF: Idiopathic Pulmonary Fibrosis; LC: Lung Cancer; COPD: Chronic Obstructive Pulmonary Disease; PC: Prostate Cancer.

Table 4. ENPP2 DMA between more (SKHEP1; PC3) and less (HEPG2; LNCAP) invasive cell lines of PC and HCC cancer types. Adopted from (Panagopoulou, Fanidis et al. 2021).

HCC							
cg00320790	0.808	0.924	0.115	Up	Body	3.1×10^{-2}	
cg09444531	0.353	0.765	0.411	Up	Body	6.5×10^{-3}	
cg20048037	0.564	0.859	0.295	Up	Body	2.4×10^{-2}	
cg07236691	0.689	0.126	-0.563	Down	Body	3.6×10^{-3}	
cg04452959	0.339	0.784	0.444	Up	TSS200	5.1×10^{-2}	
cg02156680	0.472	0.868	0.396	Up	TSS1500	3.2×10^{-3}	
cg06998282	0.637	0.938	0.302	Up	TSS1500	1.8×10^{-4}	
cg02534163	0.710	0.967	0.257	Up	1st Exon	2.7×10^{-2}	
PC							
cg00320790	0.574	0.900	0.326	Up	Body	2.3×10^{-3}	
cg07236691	0.778	0.824	0.046	Up	Body	5.6×10^{-2}	
cg09444531	0.251	0.585	0.333	Up	Body	4.3×10^{-3}	
cg20048037	0.370	0.671	0.301	Up	Body	8.8×10^{-3}	
cg20162626	0.190	0.552	0.362	Up	Body	1.7×10^{-2}	
cg26078665	0.662	0.772	0.110	Up	Body	2.3×10^{-3}	
cg02156680	0.725	0.377	-0.348	Down	TSS1500	2.5×10^{-3}	
cg02534163	0.772	0.951	0.179	Up	1st Exon	2.9×10^{-3}	

* Mean β (M β) value 1 represents methylation in less invasive cell lines and * Mean β (M β) value 2 methylation in more invasive; # $\Delta\beta$ value: (Mean β value 2-Mean β value 1). Abbreviations: PC: Prostate Cancer, HCC: Hepatocellular Carcinoma, PA: Promoter Associated, TSS: Transcription Start Site.

invasive cell lines. Last, and to investigate drug-reversibility of *ENPP2* epigenetic modification, we reanalyzed data from a colon cancer cell line treated with 5-aza-20-deoxycytidine (Seelan, Mukhopadhyay et al. 2018). Results validated our expectations and methylation levels decreased in all 14 *ENPP2* DMCs (Table S7).

Collectively, the above results support the existence of a deregulated *ENPP2* methylation pattern which in turn suggests down-regulation of its transcription levels during cancer. Increased CG methylation in more aggressive cell lines further implicates DNA methylation in cancer progression.

3.3.2 *ENPP2* expression and methylation in prostate cancer

Subsequently, we set to investigate the impact of *ENPP2* methylation changes on *ENPP2* mRNA levels. To begin with, we assessed the probability of DNA methylation affecting TF-DNA interaction. Using PROMO computational tool we predicted that 40 TFs can potentially bind on or in close proximity to multiple *ENPP2* promoter CGs (Figure 12). Four of the latter have also been found differentially methylated. Afterwards, we set to examine *ENPP2* expression and methylation in prostate (PC), lung (LC) cancer and hepatocellular carcinoma (HCC) samples. TCGA was selected as a source database due to its large cohort sizes, measurement of multiple omics modalities per sample and the provision of rich clinical and demographic metadata.



Figure 12 Transcription factors that are predicted to bind at *ENPP2* promoter CG sites. Adopted from (Panagopoulou, Fanidis et al. 2021).

To examine *ENPP2* methylation and expression during PC, we compared 235 tumor and 35 healthy adjacent tissue samples (Table 5). A dozen DMCs were found between phenotypes, while ten of them were also highlighted during up-stream analyses (Table 3). More specifically, all cancer samples' TSS and 1st exon CGs had an increased number of attached methyl groups, in contrast to the decreased methylation observed in half of the gene body CGs. Examination of methylation levels along with several cohort characteristics revealed that methylation of cg02534163 (1st exon), cg02709432 (TSS200) and cg23725583 (gene body) was positively correlated to tumor size ($p=0.032$). In addition, resistance to pharmacotherapy was related to increased methylation of cg01243251 ($p=0.023$) located at the gene's body. On the other hand, *ENPP2* expression in the same PC samples was found decreased compared to their healthy counterparts ($\text{LogFC} = -0.379$, $\text{FDR} = 3.70 \times 10^{-2}$). Spearman's correlation of gene expression and methylation at gene and isoform levels was then employed and linear models were also fit in an attempt to connect the two modalities (Figure 13A; Table 6). Despite the small effect size, cg02156680, cg02709432, cg06998282, cg14409958 (TSS CGs) and cg02534163 (1st exon CGs) methylation was negatively correlated to gene expression during PC. No significant correlation was established in control samples. At the isoform level, ucc003yot.1, uc003yos.1 and uc003yor.1 expression showed statistically significant correlation to multiple CGs with a mostly negative rho statistic.

Table 5. *ENPP2* differential methylation and expression analysis from TCGA-originating samples. Adopted from (Panagopoulou, Fanidis et al. 2021).

CG ID	M β Value 1 *	M β Value 2 *	$\Delta\beta$ Value #	Regulation	Gene Region	FDR
cg07236691	0.568	0.730	0.162	Up	Body	1.44×10^{-6}
cg09444531	0.697	0.735	0.038	Up	Body	1.09×10^{-3}
cg20048037	0.885	0.815	-0.070	Down	Body	4.22×10^{-3}
cg20162626	0.796	0.650	-0.146	Down	Body	1.44×10^{-6}
cg23725583	0.858	0.877	0.019	Up	Body	9.05×10^{-3}
cg01243251	0.933	0.904	-0.029	Down	Body	1.44×10^{-6}
cg14409958	0.251	0.590	0.339	Up	TSS1500	1.44×10^{-6}
cg02156680	0.156	0.543	0.386	Up	TSS1500	1.44×10^{-6}
cg06998282	0.212	0.645	0.433	Up	TSS1500	1.44×10^{-6}
cg02709432	0.213	0.586	0.372	Up	TSS200	1.44×10^{-6}
cg04452959	0.130	0.466	0.335	Up	TSS200	1.44×10^{-6}
cg02534163	0.128	0.565	0.436	Up	1st Exon	1.44×10^{-6}

$\Delta\beta$ value: (Mean β value 2 * cancer - Mean β value 1 * normal). Abbreviations: PC: Prostate Cancer, FC: Fold Changes.

Table 6. Correlation between *ENPP2* expression and methylation in PC, LC, HCC TCGA samples. Adopted from (Panagopoulou, Fanidis et al. 2021).

Sample Type	CG	Gene Region	Rho	FDR	Correlation	Coefficient p -Value
PC						
Tumor	cg06998282	TSS1500	-0.253	1.22×10^{-3}	Negative	1.55×10^{-2}
	cg02156680	TSS1500	-0.212	2.88×10^{-3}	Negative	4.43×10^{-2}
	cg14409958	TSS1500	-0.221	2.18×10^{-3}	Negative	3.90×10^{-2}
	cg02709432	TSS200	-0.176	1.30×10^{-2}	Negative	4.39×10^{-2}
	cg02534163	1st Exon	-0.226	2.18×10^{-3}	Negative	1.30×10^{-2}
LC						
Tumor	cg06998282	TSS1500	-0.142	>0.05	Negative	3.74×10^{-3}
	cg14409958	TSS1500	-0.213	2.19×10^{-2}	Negative	1.05×10^{-4}
Control	cg07236691	Body	-0.564	>0.05	Negative	2.31×10^{-2}
HCC						
Tumor	cg00320790	Body	0.297	9.46×10^{-6}	Positive	6.29×10^{-5}
	cg01243251	Body	0.247	2.94×10^{-4}	Positive	6.90×10^{-4}
	cg07236691	Body	0.239	4.06×10^{-4}	Positive	2.28×10^{-4}
	cg09444531	Body	0.395	1.22×10^{-9}	Positive	1.53×10^{-7}
	cg20048037	Body	0.436	8.81×10^{-12}	Positive	1.02×10^{-8}
	cg20162626	Body	0.473	0.00×10^0	Positive	2.10×10^{-9}
	cg06998282	TSS1500	-0.137	>0.05	Negative	4.55×10^{-2}
	Control	cg20162626	Body	0.422	2.63×10^{-2}	Positive
cg23725583		Body	-0.35	4.74×10^{-2}	Negative	2.61×10^{-2}
cg02709432		TSS200	-0.361	4.56×10^{-2}	Negative	3.25×10^{-2}
cg04452959		TSS200	-0.411	2.63×10^{-2}	Negative	2.87×10^{-2}
cg06998282		TSS1500	-0.464	1.60×10^{-2}	Negative	4.53×10^{-3}
cg02156680		TSS1500	-0.393	2.99×10^{-2}	Negative	3.52×10^{-2}

3.3.3 *ENPP2* expression and methylation in lung cancer

Regarding LC, we examined 212 adenocarcinoma and 15 healthy lung samples. Differential methylation analysis identified 8 DMCs, four of which have also been found during our initial GEO dataset re-analysis (Table 7). Specifically, 3 TSS and a single 1st exon CG were hyper-methylated during cancer. The opposite pattern was observed for the four gene body DMCs. When correlated to available sample metadata, methylation of TSS-related cg14409958 related to advanced cancer ($p=0.035$). DEA demonstrated lower *ENPP2* transcript levels in LC samples compared to the respective controls (LogFC= -1.285, FDR= $< 1.00 \times 10^{-2}$), similarly to PC cases. Spearman's correlation and linear models fit were once again exploited for uncovering expression-methylation relationships. Although characterized by a small effect size, *ENPP2* transcription was negatively correlated with cg14409958 methylation during LC (Figure 13B and Table 6). No significant correlation was found in control samples. Last, no statistically significant relationships were established between methylation and expression of *ENPP2* isoforms, although large effect sizes and significant linear models suggested otherwise.

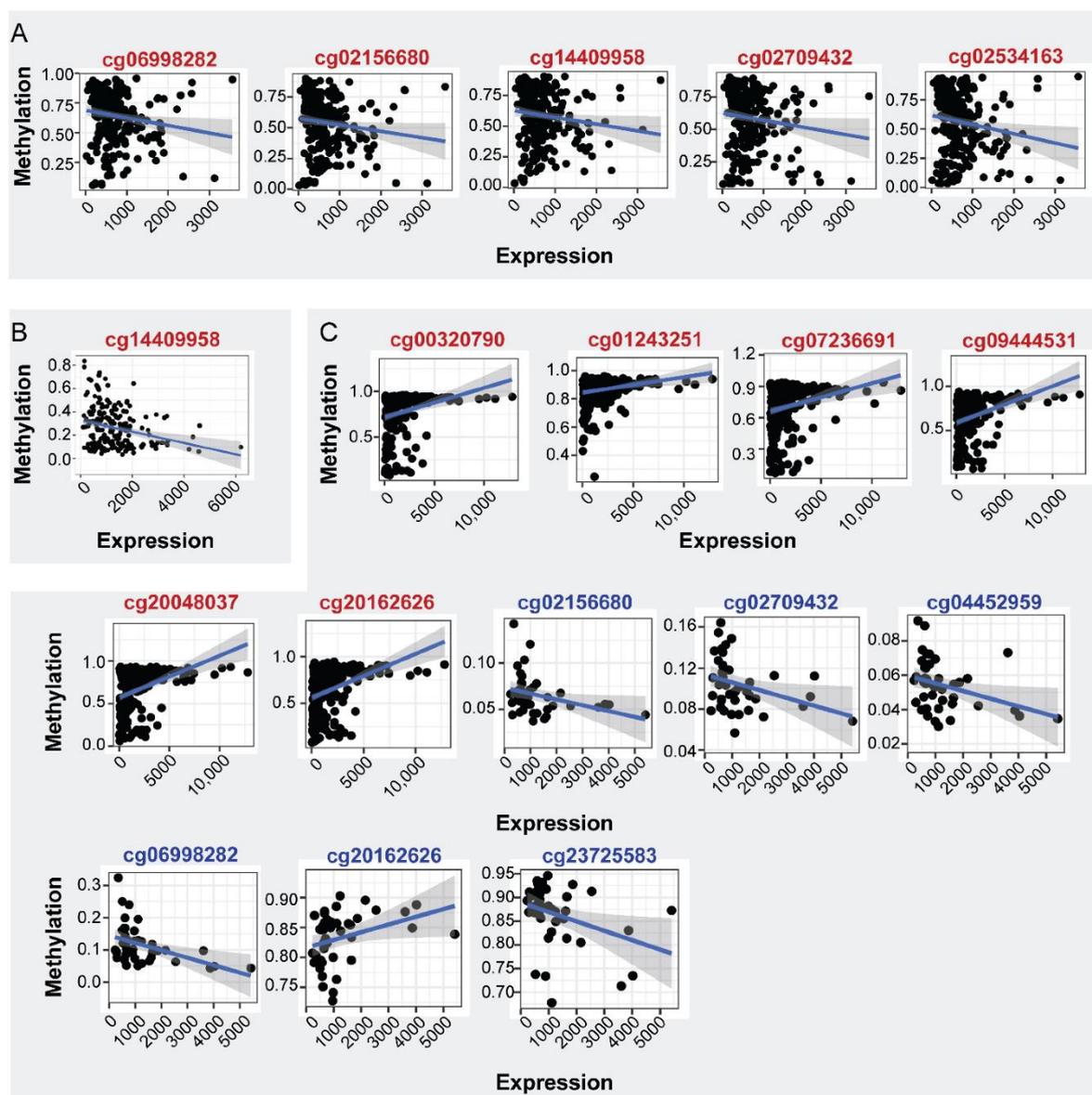


Figure 13 Significant relationships between *ENPP2* transcription and methylation. **A.** Prostate cancer. **B.** Lung cancer. **C.** Hepatocellular carcinoma. Red and blue color marks tumor and control samples, respectively. Adopted from (Panagopoulou, Fanidis et al. 2021).

Table 7. Differential methylation analysis for *ENPP2* CGs between LC and control samples. Adopted from (Panagopoulou, Fanidis et al. 2021).

CG ID	M β Value 1 *	M β Value 2 *	$\Delta\beta$ Value #	Regulation	Gene Region	FDR
cg20162626	0.750	0.636	-0.114	Down	Body	1.19×10^{-2}
cg20048037	0.722	0.624	-0.098	Down	Body	5.18×10^{-2}
cg07236691	0.561	0.547	-0.014	Down	Body	2.95×10^{-4}
cg09444531	0.730	0.630	-0.100	Down	Body	2.95×10^{-4}
cg04452959	0.071	0.138	0.066	Up	TSS200	7.19×10^{-3}
cg14409958	0.105	0.262	0.158	Up	TSS1500	2.95×10^{-4}
cg06998282	0.096	0.230	0.134	Up	TSS1500	1.05×10^{-2}
cg02534163	0.109	0.255	0.145	Up	1st Exon	8.19×10^{-4}

M β Value 1 *: Mean β value normal, M β Value 2 *: Mean β value cancer # $\Delta\beta$ value: (Mean β value 2 * cancer - Mean β value 1 * normal).
Abbreviations: LC: Lung Cancer, FC: Fold Changes.

3.3.4 *ENPP2* expression and methylation in hepatocellular carcinoma

With regard to HCC, we retrieved and examined 241 HCC and 42 control liver samples. Differential methylation analysis identified 13 DMCs, 12 of which were shared with the HCC GEO datasets analyzed earlier (Table 8). Once again, gene body sites were down-methylated during cancer, in contrast to the TSS and 1st exon GCs. Examination of clinical characteristics suggested that younger HCC patients had more methyl groups attached to 7 of their CGs (cg00320790, cg01243251, cg07236691, cg09444531, cg20048037, cg20162626 and cg2372558) compared to older ones (all $p < 0.001$). On the other hand, cg01243251 methylation was negatively correlated with control samples age ($p = 0.0037$). Methylation of most CGs was also correlated to patients' sex, with female individuals presenting a heavier decoration than males (all $p < 0.05$). We observed the same phenomenon in just a single CG during homeostasis (cg20048037; $p = 0.033$). Interestingly, tumor invasiveness was positively related to cg04452959 methylation ($p = 0.044$). In contrast to PC and LC, *ENPP2* expression was increased during HCC (LogFC= 0.710, FDR= 1.00×10^{-2}). Subsequently, methylation of 6 gene body CGs was positively associated with *ENPP2* expression (Figure 13C and Table 6). On the contrary, almost all statistically significant steady state relationships followed the opposite pattern. Correlation of methylation and isoform level expression produced similar results.

Table 8. Differential methylation analysis for *ENPP2* CGs between LC and control samples. Adopted from (Panagopoulou, Fanidis et al. 2021).

CG ID	Mβ Value 1 *	Mβ Value 2 *	Δβ Value #	Regulation	Gene Region	FDR
cg00320790	0.929	0.798	-0.131	Down	Body	3.90×10^{-5}
cg01243251	0.917	0.872	-0.046	Down	Body	1.99×10^{-3}
cg07236691	0.878	0.734	-0.143	Down	Body	1.30×10^{-5}
cg09444531	0.850	0.689	-0.161	Down	Body	1.30×10^{-5}
cg20048037	0.832	0.684	-0.148	Down	Body	4.04×10^{-2}
cg20162626	0.833	0.671	-0.161	Down	Body	3.75×10^{-4}
cg23725583	0.864	0.765	-0.099	Down	Body	1.50×10^{-2}
cg04452959	0.054	0.180	0.126	Up	TSS200	4.46×10^{-2}
cg02709432	0.104	0.264	0.160	Up	TSS200	2.83×10^{-4}
cg02156680	0.065	0.210	0.144	Up	TSS1500	3.76×10^{-2}
cg06998282	0.116	0.300	0.184	Up	TSS1500	1.86×10^{-2}
cg14409958	0.340	0.463	0.123	Up	TSS1500	1.86×10^{-2}
cg02534163	0.070	0.247	0.177	Up	1st Exon	3.15×10^{-4}

Mβ Value 1 *: Mean β value normal, Mβ Value 2 *: Mean β value cancer # Δβ value: (Mean β value cancer-Mean β value normal).
Abbreviations: LC: Lung Cancer, FC: Fold Changes.

3.3.5 *ENPP2* expression and methylation in UALCAN database data

To replicate our findings in an independent sample set, we re-analyzed UALCAN database-retrieved data. In accordance to TCGA findings, *ENPP2* expression was decreased in PC/LC and increased in HCC cases. Moreover, differential protein levels analysis in LC revealed reduction in primary tumors relative to control cases. In agreement to our previous observations, *ENPP2* was hyper-methylated in all three cancer types compared to respective controls.

To conclude with, data from various sources converge on the existence of both *ENPP2* differential methylation and expression in cancer affected tissues. Methylation in PC and LC follows the same hyper-methylation motif of TSS and 1st exon CGs. On the other hand, addition of methyl groups during HCC seems to obey a cancer type-specific mechanism. Although, pending wet lab validation studies, our *in silico* analyses suggest the existence of a potentially drugable regulatory link between *ENPP2* expression and methylation.

3.4 *ENPP2* methylation as a liquid biomarker of breast cancer

Motivated by our findings in PC, LC and HCC, we next decided to examine *ENPP2* methylation and expression in breast cancer, the most common cancer type between female individuals (Łukasiewicz, Czezelewski et al. 2021).

3.4.1 *ENPP2* methylation and expression in breast cancer

To begin with, we analyzed data from 520 primary/metastatic breast cancer (BrCa) and 185 control tissue samples originating from five GEO datasets and the TCGA database. Differential methylation analysis revealed that all TSS and 1st exon CGs (cg04452959, cg02709432, cg02156680, cg06998282 and cg02534163) were hyper-methylated during cancer (Table 9). The same pattern was followed by three gene body CGs (cg09444531, cg26078665, cg23725583), while cg00320790 and cg20048037 were hypo-methylated in BrCa cases. DEA of the above TCGA data (302 BrCa and 76 normal tissue samples) demonstrated a decrease in *ENPP2* transcript levels in cancer-affected tissue samples (FC=-5.15, FDR=3.96×10⁻⁶⁶) (Table 10).

Table 9. *ENPP2* DMCs between BrCa and control tissue samples. Adopted from (Panagopoulou, Drosouni et al. 2022).

CG ID	Mβ Value Normal	Mβ Value BrCa	Δβ Value	Methylation in BrCa	Gene Region	FDR
Normal breast tissue vs. BrCa						
cg00320790	0.96	0.95	0.01	Down	Body	5.97 × 10 ⁻⁴
cg20048037	0.92	0.87	0.05	Down	Body	1.13 × 10 ⁻¹²
cg09444531	0.77	0.79	-0.02	Up	Body	5.16 × 10 ⁻³
cg26078665	0.77	0.84	-0.07	Up	Body	7.32 × 10 ⁻¹⁴
cg23725583	0.85	0.92	-0.06	Up	Body	1.03 × 10 ⁻¹⁵
cg02534163	0.06	0.53	-0.47	Up	1st Exon	3.15 × 10 ⁻⁹¹
cg04452959	0.03	0.44	-0.41	Up	TSS200	4.56 × 10 ⁻⁸⁰
cg02709432	0.09	0.57	-0.48	Up	TSS200	6.71 × 10 ⁻⁷³
cg02156680	0.04	0.44	-0.39	Up	TSS1500	9.18 × 10 ⁻⁷²
cg06998282	0.09	0.62	-0.53	Up	TSS1500	9.54 × 10 ⁻⁷⁶
Primary vs. Metastatic BrCa						
cg20048037	0.87	0.82	0.06	Down	Body	3.99 × 10 ⁻²
cg09444531	0.78	0.71	0.06	Down	Body	3.77 × 10 ⁻²
cg26078665	0.86	0.79	0.07	Down	Body	9.09 × 10 ⁻⁴
cg23725583	0.92	0.88	0.04	Down	Body	2.56 × 10 ⁻²
cg02534163	0.55	0.74	-0.19	Up	1st Exon	1.26 × 10 ⁻⁴
cg06998282	0.64	0.79	-0.15	Up	TSS1500	2.28 × 10 ⁻³
Normal breast vs. stage I BrCa						
cg20048037	0.92	0.89	0.03	Down	Body	4.35 × 10 ⁻⁴
cg09444531	0.77	0.80	-0.03	Up	Body	9.27 × 10 ⁻³
cg26078665	0.78	0.86	-0.08	Up	Body	1.07 × 10 ⁻⁷
cg23725583	0.86	0.93	-0.07	Up	Body	1.74 × 10 ⁻⁸
cg02534163	0.06	0.55	-0.48	Up	1st Exon	2.45 × 10 ⁻⁴⁹
cg04452959	0.04	0.47	-0.43	Up	TSS200	5.53 × 10 ⁻³⁸
cg02709432	0.10	0.61	-0.51	Up	TSS200	1.14 × 10 ⁻³⁸
cg02156680	0.05	0.47	-0.43	Up	TSS1500	5.59 × 10 ⁻³⁹
cg06998282	0.10	0.66	-0.56	Up	TSS1500	2.93 × 10 ⁻³⁵
Early vs. Advanced BrCa						
cg01243251	0.94	0.92	0.014	Down	Body	3.10 × 10 ⁻²
cg20162626	0.75	0.69	0.051	Down	Body	3.12 × 10 ⁻³

Table 10. *ENPP2* DE between cancer and control TCGA samples (|FC| ≤ 1.2; FDR < 0.05). Adopted from (Panagopoulou, Drosouni et al. 2022).

Compared Breast Groups	Fold Change	p-Value	FDR
Cancer_vs_Normal	-5.15	1.18 × 10 ⁻⁶⁷	3.96 × 10 ⁻⁶⁶
StageI_vs_Normal	-5.46	6.28 × 10 ⁻⁵⁴	3.43 × 10 ⁻⁵²
Advanced_vs_Early	1.20	1.23 × 10 ⁻²	9.41 × 10 ⁻²

Last, correlation of *ENPP2* expression and methylation status was recruited to reveal any potential connection between the two modalities (Figure 14A, Table 11). Regarding BrCa cases

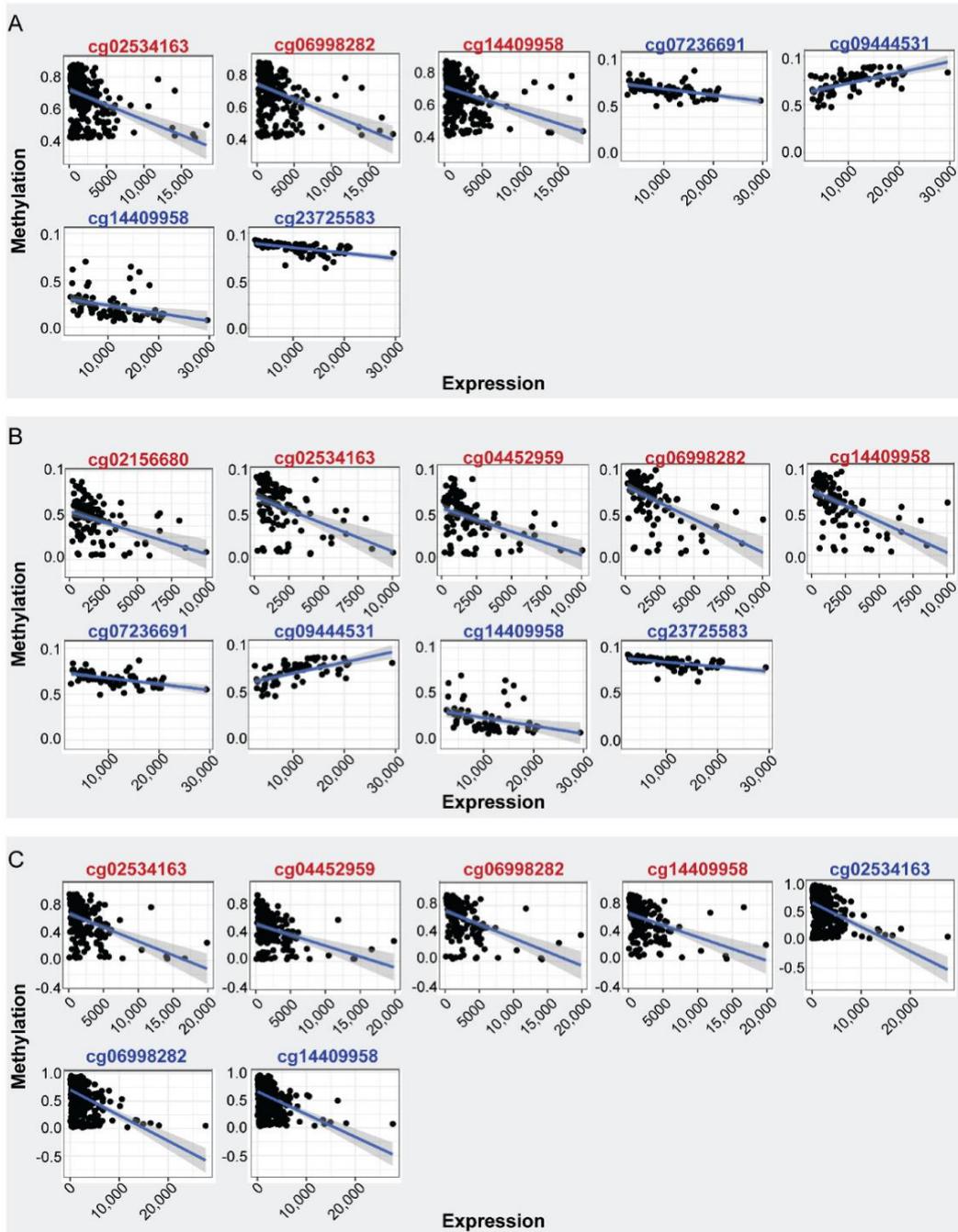


Figure 14 Relationship of *ENPP2* transcription and methylation in breast cancer samples. Depicted are cases of significant correlation between CG methylation and *ENPP2* expression (Spearman's correlation; $|\rho| \geq 0.4$; $FDR < 0.05$). **A.** Breast cancer (red) and control samples (blue). **B.** Breast cancer stage I (red) and control samples (blue). **C.** Advanced (red) and early (blue) breast cancer samples. Adopted from (Panagopoulou, Drosouni et al. 2022).

transcript levels were negatively correlated with promoter-associated CGs (cg02534163, cg06998282, cg14409958) and no significant relationship was established with gene body methylation sites. Analysis of normal samples highlighted a negative correlation between *ENPP2* transcription and two gene body CGs (cg07236691, cg23725583). Intriguingly, and a

single TSS dinucleotide which was also among the significant tumor correlations: cg14409958. A sole positive correlation was revealed for cg14409958 of the gene body. All together, the above results suggest that *ENPP2* genomic site does undergo differential methylation during BrCa, with hypermethylation of promoter-associated sites reducing the genes transcription rates.

Table 11. Correlation between *ENPP2* transcription and methylation for different BrCa stages and control samples. Adopted form (Panagoulou, Drosouni et al. 2022).

Tissue	CG	BrCa vs. Normal		FDR	Correlation
		Gene Region	Rho		
BrCa	cg02534163	First Exon	-0.40	1.18×10^{-12}	Negative
	cg06998282	TSS1500	-0.42	1.53×10^{-13}	Negative
	cg14409958	TSS1500	-0.42	2.10×10^{-13}	Negative
Normal	cg09444531	Body	0.62	4.77×10^{-06}	Positive
	cg23725583	Body	-0.70	3.02×10^{-11}	Negative
	cg07236691	Body	-0.55	1.27×10^{-08}	Negative
	cg14409958	TSS1500	-0.52	9.53×10^{-07}	Negative
Stage I BrCa vs. Normal					
Stage I	cg02534163	First Exon	-0.46	1.54×10^{-06}	Negative
	cg04452959	TSS200	-0.44	3.74×10^{-06}	Negative
	cg14409958	TSS1500	-0.64	3.29×10^{-13}	Negative
	cg02156680	TSS1500	-0.42	1.28×10^{-05}	Negative
	cg06998282	TSS1500	-0.63	6.04×10^{-13}	Negative
Normal	cg09444531	Body	0.64	3.25×10^{-08}	Positive
	cg23725583	Body	-0.66	1.36×10^{-08}	Negative
	cg07236691	Body	-0.53	1.40×10^{-05}	Negative
	cg14409958	TSS1500	-0.51	4.25×10^{-05}	Negative
Early vs. Advanced BrCa					
Early	cg02534163	First Exon	-0.42	2.56×10^{-23}	Negative
	cg06998282	TSS1500	-0.47	4.588×10^{-29}	Negative
	cg14409958	TSS1500	-0.46	5.37×10^{-28}	Negative
Advanced	cg02534163	First Exon	-0.43	1.23×10^{-09}	Negative
	cg04452959	TSS200	-0.41	5.47×10^{-09}	Negative
	cg14409958	TSS1500	-0.48	4.72×10^{-12}	Negative
	cg06998282	TSS1500	-0.49	4.18×10^{-12}	Negative

Abbreviations: FDR: false discovery rate, TSS: Transcription Start Site.

3.4.2 *ENPP2* methylation and expression between breast cancer stages and types

As breast cancer is considered curable if detected prior to any metastatic event (Harbeck, Penault-Llorca et al. 2019), discovery of an accurate and sensitive detection strategy is of utmost importance. Initially, before investigating the diagnostic potential of *ENPP2* methylation, we wanted to examine whether methylation changes can indeed be detected early on BrCa progress. Towards this purpose we re-analyzed raw data from 132 primary and 31 distant metastatic cancer cases. Four gene body CGs were found hypo-methylated in metastatic cancer while two others, one in the 1st exon and the other in the TSS area, were conversely affected (Table 9), proposing the implication of *ENPP2* methylation in the BrCa progress and/or invasiveness.

To replicate and extend the above findings, we subsequently analyzed TCGA data from stage-I BrCa and healthy tissue samples. Differential methylation analysis based on 136 stage-I and 111 normal specimens identified 9 DMCs, with five of the hyper-methylated ones located at promoter-associated regions and another three at the *ENPP2* body (Table 9). Consisting the sole divergence from the general pattern, cg20048037 was found hypo-methylated during BrCa onset. In addition, DEA in 111 stage-I and 66 normal tissue areas revealed once again the *ENPP2* transcription downregulation (FC=-5.46, FDR=3.43×10⁻⁵²) (Table 10). When we examined expression and methylation levels of stage-I BrCa samples, we observed exclusively negative relationships (Table 11). The same pattern was witnessed in normal-origin data implicating one TSS and two body CGs, while a positive correlation was uncovered in a single gene body instance.

Afterwards, we focused on any epigenetic regulation taking place during BrCa latter stages, using data from 521 early (stage I, II) and 221 advanced (stage III) cases. Interestingly, only two gene body DMCs were identified (Table 9). Moreover, we established no difference (FC=1.20, FDR=9.41×10⁻²) in *ENPP2* transcript levels between 519 early and 191 advanced BrCa examined individuals (Table 10). Inability of methylation and expression values to separate samples by phenotype, as revealed through dimensionality reduction methods applied on data of both modalities (Figure S2), further validates/supports the lack of significant findings. Nevertheless, irrespective of these negative findings, significant correlations were identified between *ENPP2* transcription and methylation of three TSS and a 1st exon CG in advanced cancer cases (Figure 14C; Table 11). Negative relationships in three out of the four locations mentioned above were reported in early BrCa samples, too.

In order to examine differences between invasive cancer types, we retrieved TCGA data from 473 invasive ductal and 186 invasive lobular suffering individuals. Differential methylation analysis revealed only three statistically significant DMCs, which unfortunately had very small effect sizes (Table 4).

Table 12. *ENPP2* CGs differentially methylated between ductal and lobular BrCa. Adopted from (Panagopoulou, Drosouni et al. 2022).

CG ID	Mβ Value Ductal Cancer	Mβ Value Lobular Cancer	Δβ Value	Methylation in Ductal Cancer	Gene Region	FDR
cg01243251	0.94	0.93	0.01	Up	Body	1.53 × 10 ⁻²
cg20048037	0.86	0.90	-0.04	Down	Body	1.11 × 10 ⁻²
cg02156680	0.47	0.52	-0.04	Down	TSS1500	2.52 × 10 ⁻²

Abbreviations: BrCa: breast cancer, DMCs: differentially methylated CpGs, FDR: false discovery rate, Mβ Value: mean β value, Δβ Value: difference between mean values.

Last, UALCAN database was used as an independent data source for validation of all aforementioned findings. Indeed, not only promoter-associated regions had increased levels of methylation compared to normal tissue samples (Figure 15A), but *ENPP2* expression was found downregulated in cases of primary tumor (Figure 15B). In addition, protein levels were also reduced in the same tumor samples respective to their control counterparts (Figure 15C).

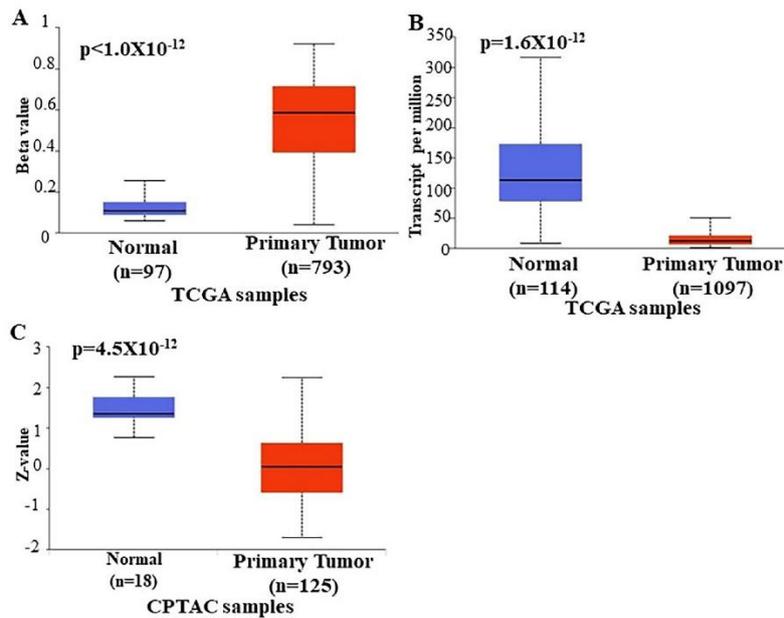


Figure 15 Analysis of *ENPP2* transcription and its regulation from UAFLAN database. A. *ENPP2* promoter methylation, B. mRNA expression and C. protein expression from BrCa and control samples recovered from the UAFLAN platform. CPTAC: clinical proteomic tumor analysis consortium. Adopted from (Panagopoulou, Drosouni et al. 2022).

Collectively, data analyses strongly indicate that *ENPP2* methylation does change alongside cancer establishment, with increased accumulation of methyl groups on promoter-related CG sites causing the reduction of *ENPP2* transcript and protein levels. Nevertheless, once cancer reaches a severe status, no high effect changes can be detected in neither *ENPP2* expression nor methylation in respect to earlier cancer stages.

3.4.3 *ENPP2* methylation and expression between breast cancer types

Reduced and even minimal invasiveness is key for the wide application of a diagnostic test. As far as cancer diagnosis is concerned, several studies have been able to measure methylation via liquid biopsies and then relate it to cancer evolution (Panagopoulou, Karaglani et al. 2019, Panagopoulou, Cheretaki et al. 2021, Ren, Lu et al. 2022). Motivated by the previously performed investigations, we wanted to discover if *ENPP2* methylation does offer the same diagnostic capabilities.

Working towards that goal, we re-analyzed a GEO-deposited dataset (GSE1222126) of circulating cell free DNA (ccfDNA) revealing eight DMCs between cancer and control samples (Table 13). In detail, all promoter-associated CGs were found hyper-methylated with big effect sizes, while the opposite pattern was detected in a single gene body CG. Most significantly, half of the aforementioned CGs located in promoter-associated regions (cg02534163, cg04452959, cg02156680 and cg06998282), have also been reported as DMCs in the BrCa_vs_normal analyses presented above (Table 9). Thus, these results propose that ccfDNA may indeed work as a window to *ENPP2* methylation status in cancer-affected tissue sites.

3.4.4 *ENPP2* methylation as measured in ccfDNA

In an attempt to validate our *in silico* findings and support the clinical validity of measuring *ENPP2* methylation in liquid biopsy samples, we have used qMSP to target cg02534163, a 1st exon CG identified as a DMC in computational analyses of data from both lung tissue and the

Table 13. Comparison of *ENPP2* methylation between BrCa and control individuals as represented in ccfDNA samples. Adopted from (Panagopoulou, Drosouni et al. 2022).

CG ID	M β Value BrCa	M β Value Normal	$\Delta\beta$ Value	Methylation in BrCa	Location	FDR
cg07236691	0.583	0.817	-0.234	Down	Body	3.04×10^{-3}
cg20048037	0.711	0.331	0.380	Up	Body	5.08×10^{-5}
cg20162626	0.489	0.814	-0.325	Down	Body	3.18×10^{-4}
cg02534163	0.802	0.206	0.596	Up	1st Exon	6.30×10^{-11}
cg04452959	0.620	0.020	0.599	Up	TS200	2.51×10^{-13}
cg02156680	0.515	0.028	0.487	Up	TSS1500	1.48×10^{-13}
cg06998282	0.772	0.057	0.715	Up	TSS1500	1.61×10^{-8}
cg14409958	0.748	0.053	0.695	Up	TSS1500	2.24×10^{-6}

Abbreviations: BrCa: breast cancer, DMCs: differentially methylated CpGs, FDR: false discovery rate, M β Value: mean β value, $\Delta\beta$ Value: difference between mean values.

periphery. More specifically, ccfDNA was collected from 52 adjuvant-treated, 19 metastatic and 15 neoadjuvant-treated BrCa suffering individuals. Twenty healthy individuals formed the negative control sample. Initially, it should be noted that methylation was detected more frequently in cancer-originating than in healthy-sourced ccfDNA (Figure 16A). Separate consideration of cancer types, detected *ENPP2* methylation in 71.1%, 73.6% and 72.1% of the adjuvant, metastatic and neoadjuvant cases, respectively. Comparison of methylation positives between groups revealed significant differences between control_vs_adjuvant, control_vs_metastatic and control_vs_neoadjuvant cases (Figure 16B), yet no difference was identified between BrCa groups *per se*.

Last, we performed differential methylation analysis between conditions. No significant difference was recorded between pathological and steady-state samples (Figure 17A), while the neoadjuvant-treated individuals had increased levels of methylated ccfDNA as compared to either control or adjuvant group (Figure 17B). Comparison of many- versus few-metastatic event cases revealed cg02534163 hyper-methylation in patients of the former category (Figure 17C).

In conclusion, the aforementioned observations support that use of cg02534163 methylation as detected in ccfDNA is a possible test for non-invasive diagnosis of early stage cancer. Further larger studies must be conducted for the independent validation and establishment of *ENPP2* methylation as a BrCa biomarker.

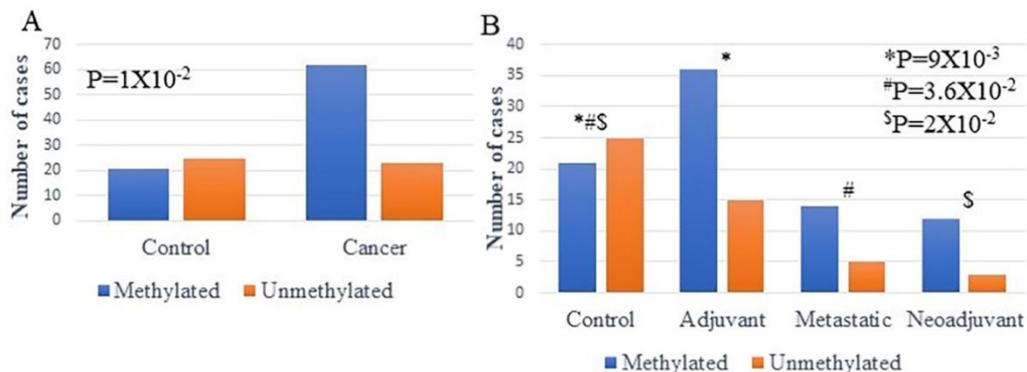


Figure 16 *ENPP2* methylation as measured from ccfDNA isolated from breast cancer affected and control individuals. A. Breast cancer and control samples. **B.** The same samples per breast cancer group. *Control related to adjuvant; # Control related to metastatic; \$ Control related to neoadjuvant. Adopted from (Panagopoulou, Drosouni et al. 2022).

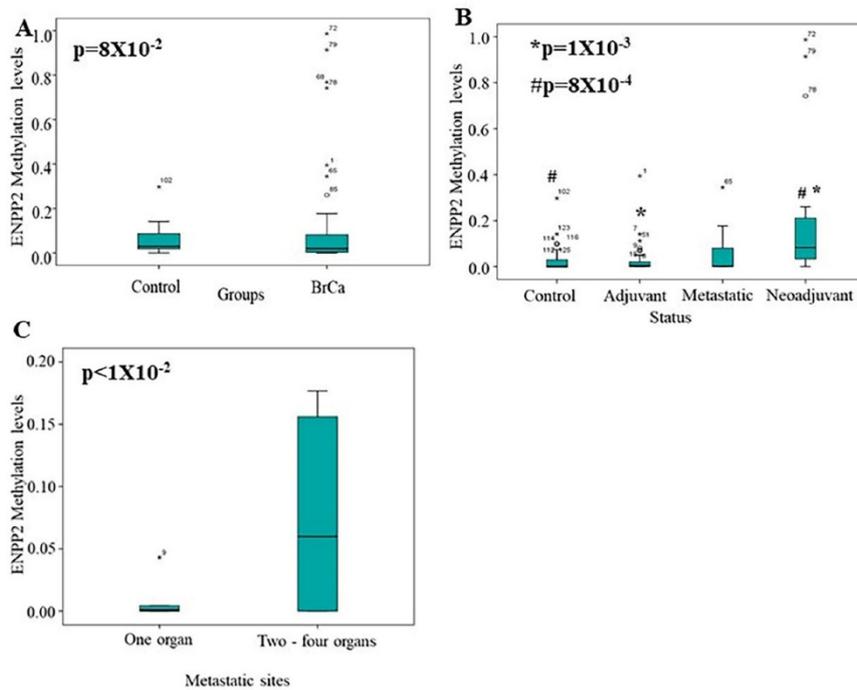


Figure 17 Differential methylation analysis between different pathology conditions and control samples. Boxplots of ccfDNA levels from (A) BrCa compared to control; (B) in each group separately; (C) BrCa patients with one or more metastatic events. *Adjuvant related to neoadjuvant; #Control related to neoadjuvant. Adopted from (Panagopoulou, Drosouni et al. 2022).

3.5 Gut, liver and lung microbiome

To examine how pulmonary homeostasis is affected by lung microbiome and if there is any interaction with microbes of both gut and liver milieu, we performed 16S rRNA sequencing from murine samples of all the above three tissues. Samples were collected from both high-fat diet (HFD) and control diet (CD) treated C57Bl6 mice, post to validation of obesity establishment in the former group.

Initially, we quality controlled and denoised FASTQ files produced by the BSRC genomics facility, using DADA2 functionalities (Callahan, McMurdie et al. 2016). Then, samples were pooled per tissue and diet prior to amplicon sequence variants (ASV) formation. ASVs were adjusted for taxa-specific 16S gene copy numbers (GCN correction). In total, we were able to detect 29 phyla, 59 classes, 130 orders, 227 families and 585 species which we used for all downstream analyses (Figure S3A).

To examine general diversity patterns, we utilized the observed number of ASVs (richness), as well as Shannon's index (biodiversity). In detail, computed richness suggested that more diverse microbial communities exist under HFD conditions in all three tissues (Figure 18A; Figure S3C). On the other hand, Shannon's index corroborates the above observation only for lung tissue samples (Figure 18B). Subsequently, to compare microbial composition between tissues and diets we calculated the Aitchison distance (β -diversity) under healthy and obese conditions. It was observed that HFD-driven obesity changes lung microbiome to a more liver-like composition (Figure 18C). Examination of phyla and families per diet and tissue suggested that obesity does increase the absolute numbers of microbial categories shared across tissues, thus potentially shaping a more homogeneous microbiome across the body (Figure 18D-E; Tables S8-9).

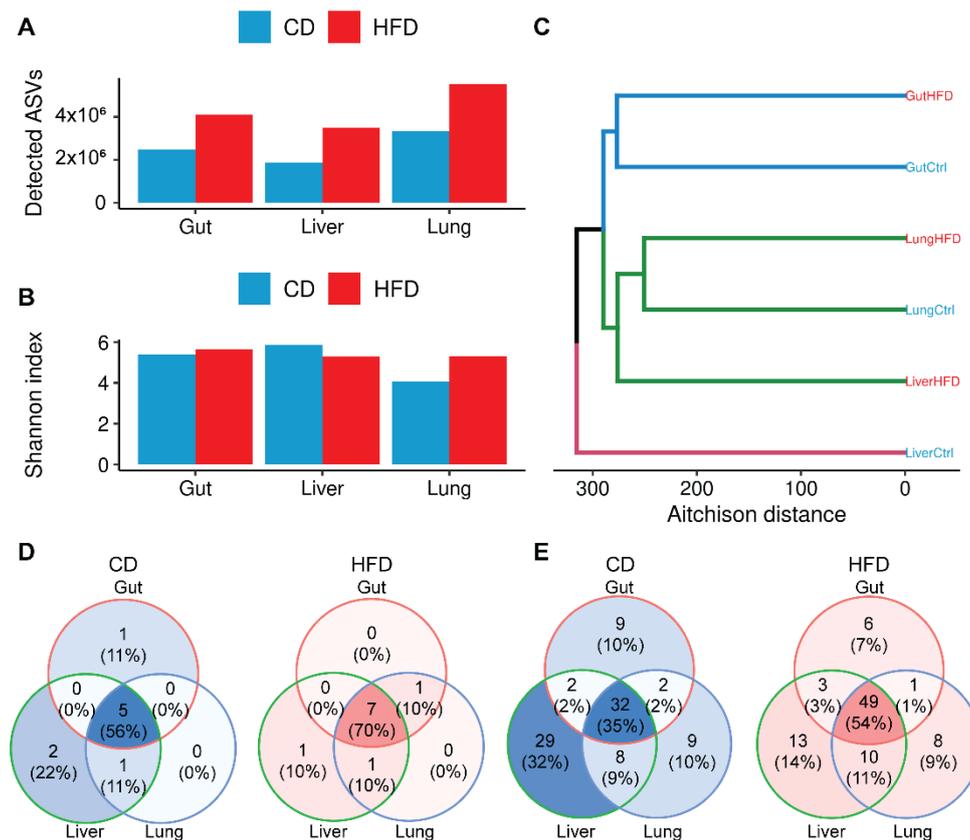


Figure 18 Microbiome of all three tissues is affected from HFD-driven obesity. **A.** A higher number of ASVs is observed under HFD-driven obesity conditions. **B.** Sample biodiversity per tissue and diet. **C.** β -diversity between tissue and diet samples. **D-E.** Common phyla and families upon CD or HFD in gut, liver, and lung. Adopted from (Galaris, Fanidis et al. 2022).

Afterwards, we attempted to dissect the above mentioned changes at both the phylum and family levels of taxa organization. After removing zero abundance phyla we discovered that all three tissues shared the same taxa, namely Proteobacteria, Actinobacteria, Bacteroidetes, Cyanobacteria and Firmicutes, irrespective of the received diet (Figure 19A-B). Nevertheless, samples hierarchical clustering suggested that HFD disrupts phyla relative abundance, rendering liver microbial community more similar to that of the lung instead of the gut. The same observation can be made from β -diversity clustering (Figure 18C), as well as from phyla intersection between conditions (Figure 18D). More specifically, Proteobacteria was the most abundant phylum in liver and lung of CD-treated mice, with Actinobacteria and Firmicutes ranking second and third, respectively (Figure 19C). Slightly different, the CD gut was primarily populated by Bacteroidetes and then by Proteobacteria and Firmicutes. In contrast, during HFD Firmicutes was the second most abundant phylum in liver and lung with Proteobacteria maintaining the first and Actinobacteria sidelined to the third position (Figure 19C). Under the same conditions, Firmicutes, Actinobacteria and Proteobacteria were the most frequently found phyla in the gut. In total, all the above data propose an expansion of Firmicutes during HFD.

Moving to the family level, Firmicutes families Staphylococcaceae, Streptococcaceae and Peptoniphilaceae families were expanding during obesity in all examined tissues (Figure 20A). Pasteurellaceae (Proteobacteria) followed the same pattern of expansion. Regarding genera of

the above three Firmicutes families (Table S10), *Staphylococcus* seems to be the most favored by HFD (Figure 20B).

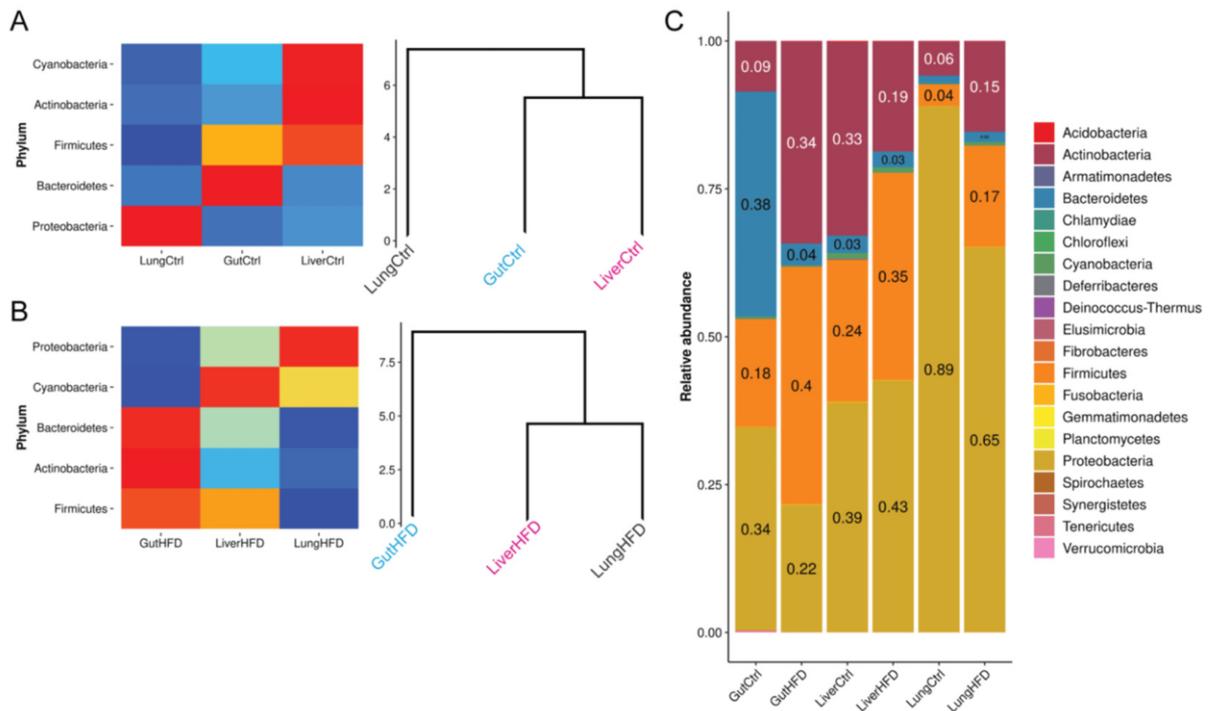


Figure 19 Relative phyla abundance in all tissues and diets. **A-B.** Heatmaps and clustering of relative abundance in CD and HFD-treated animals. Only phyla with a non-zero two-decimal places abundance were considered. Heatmaps were scaled per phylum. Manhattan distance with complete linkage was used for hierarchical clustering. **C.** Detected phyla relative abundance. Adopted from (Galaris, Fanidis et al. 2022).

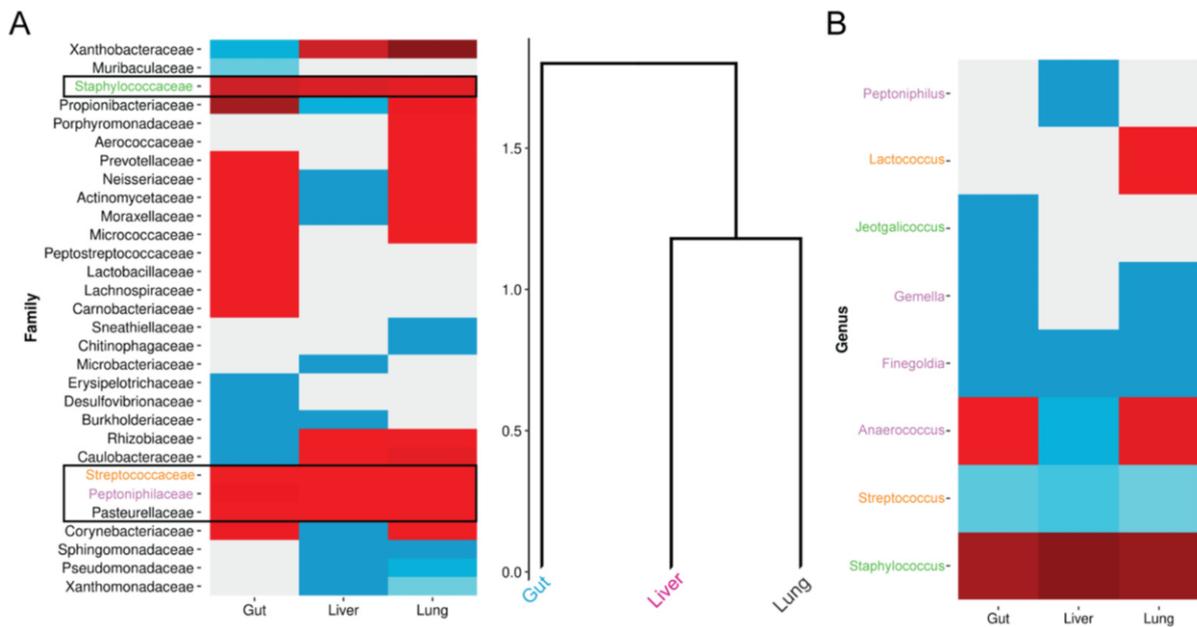


Figure 20 HFD effect on microbial families and genera. **A.** Relative abundance of families between tissues. **B.** Relative abundance of genera between tissues. Only taxa with non-zero relative abundance differences between the two diet regimes in at least one tissue were considered. Heatmaps are scaled per taxon. Manhattan distance and complete linkage were used for hierarchical clustering purposes.

Genera and their respective family are color-coded in both figure panels. Adopted from (Galaris, Fanidis et al. 2022).

Last, as the three most consistently affected Firmicutes genera, *Staphylococcus*, *Streptococcus* and *Finegoldia*, host species that have been previously witnessed to produce superantigens (Fraser and Proft 2008, Anderson, Zheng et al. 2012), we queried UniProt using the term ‘superantigen’ to uncover more such microbes. Intersection of the query results with the here detected species yielded six superantigen producers (Table S11), four of which seems to have been affected by HFD in a single organ at minimum.

In summary, all the aforementioned data propose that HFD-driven obesity does not only affect liver and gut microbiome, but also that of the lungs. Disruption of the microbiomic communities was reflected from the highest to the lowest level of taxa organization, with certain families undergoing consistent changes in their abundance. Further experimentations are absolutely necessary to support these preliminary findings, which if validated will change our perception of obesity and its effects in the respiratory system.

3.6 Lipocalin-2 in fibrosis and steady state

Motivated by a potential participation of LCN2 in the establishment of PF (see Introduction), we decided to examine the above hypothesis by means of both dry and wet laboratory processes in human and murine samples.

3.6.1 Human LCN2 expression patterns

Initially, we examined *LCN2* expression using bulk high-throughput sequencing data from Fibromine (Fanidis, Moulos et al. 2021) (Table S12). Multiple datasets mining revealed an increase in *LCN2* mRNA levels in IPF compared to control lung tissue samples (Figure 21A-B; Table S12). More importantly, expression data negatively correlated with lung function measurements (DLCO, FVC and FEV1) in IPF patients of the same datasets (Figure 21C; Figure S4A-B).

Subsequently, we set to dissect *LCN2* expression in lung cell types. To accomplish this task, we re-analyzed data from three open source scRNA-seq datasets (Table S13) (Adams, Schupp et al. 2020, Habermann, Gutierrez et al. 2020, Mayr, Simon et al. 2021). *LCN2* transcripts were detected in all cases (Figure 21D; Figure S4C-D) and expressed mainly by epithelial (goblet, ciliated, basal and club) and aberrant basaloid cells (Figures 21E; Figure S4E-F). Comparison of expression between cells originating from IPF and control individuals, demonstrated *LCN2* up-regulation in various cell types including AT1 and AT2 (Figure 21F; Figure S4G-H). Although neutrophil-produced *LCN2* has been witnessed in multiple pathologies (Hu, Li et al. 2023), low neutrophil cell numbers in the here presented scRNA-seq datasets did not enable a proper verification of literature statements. Last, to support our *in silico* observations we ELISA-measured proteins in BAL fluid from 26 IPF patients. Indeed, *LCN2* abundance negatively correlated with respiratory function measurements as suggested by *in silico* findings (Figure 22A-C).

Thus, initial data suggest that *LCN2* is over-expressed in the lungs of IPF patients, a phenomenon that seems to be related with progressive decline of lung function. Various epithelial cells, as well as the recently identified population of aberrant basaloid cells were the primary source of *LCN2* in both IPF and steady state conditions. *LCN2* can be isolated from the BAL fluid of IPF-suffering individuals, with the protein levels negatively correlating to lung function.

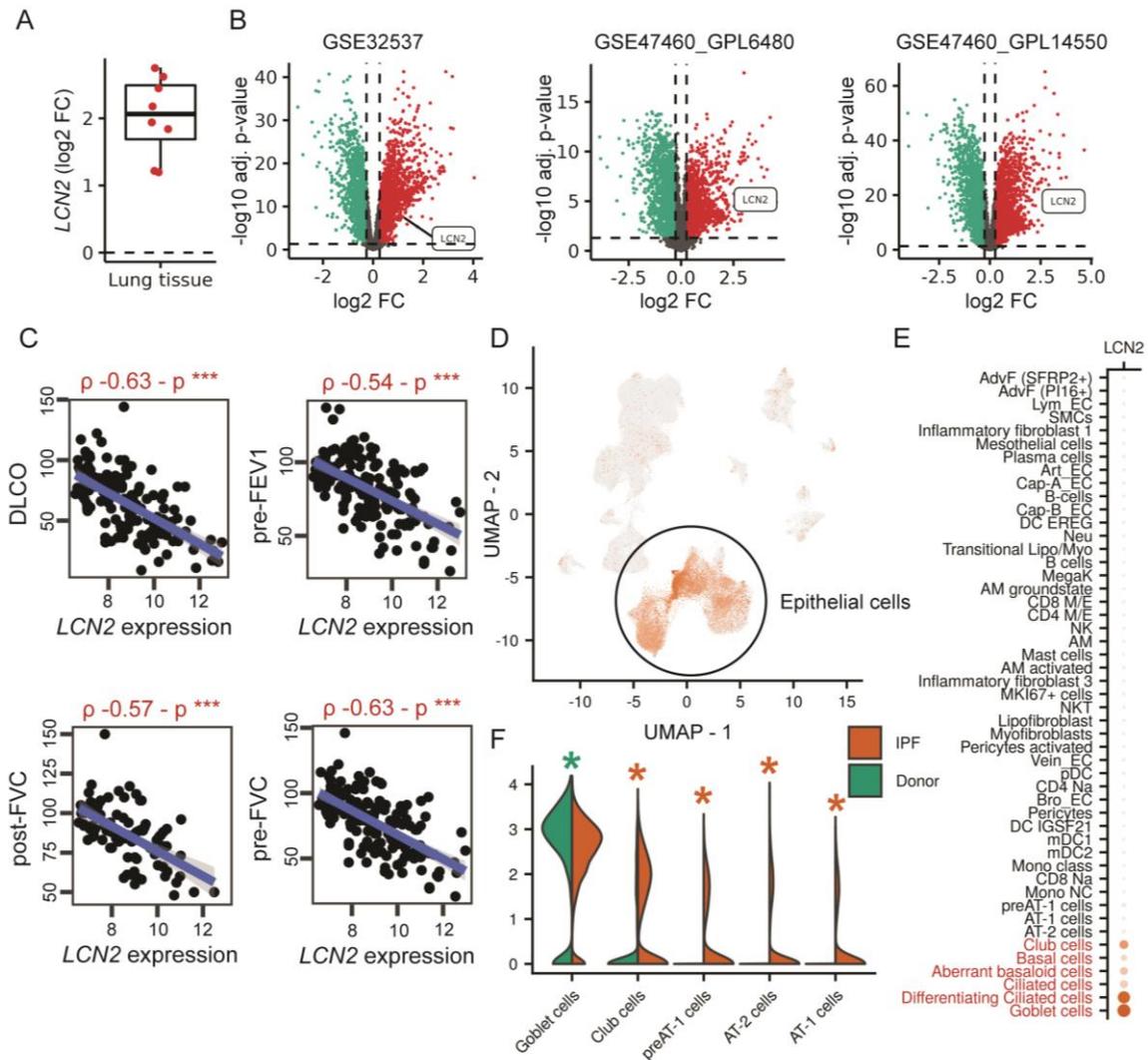


Figure 21 Increased *LCN2* expression during fibrosis is negatively correlated to lung function measurements. **A.** Differential *LCN2* expression between IPF and control tissues (Table S12). **B.** DEGs (FC > 1.2; FDR adjusted $p < 0.05$) of the largest datasets in (A) examining 115/44, 28/15, 84/75 IPF patients and controls, respectively. **C.** Spearman's correlation between *LCN2* transcription and lung function from the GSE47460_GPL14550 dataset (** $p < 0.01$). **D.** *LCN2* is mainly expressed by pulmonary epithelial cells. **E.** Dot plot of the dataset in (D) iconizing *LCN2* as a marker gene of six cell types in red font (Wilcoxon rank-sum; FC > 1.2 and Bonferroni adjusted $p < 0.05$). **F.** Differential expression between cell groups and phenotypes (IPF vs. control origin) (Wilcoxon rank-sum test; *FC ≥ 1.2 ; Bonferroni-corrected $p < 0.05$; *upregulated in IPF; *downregulated in IPF). Adopted from (Galaris, Fanidis et al. 2023).

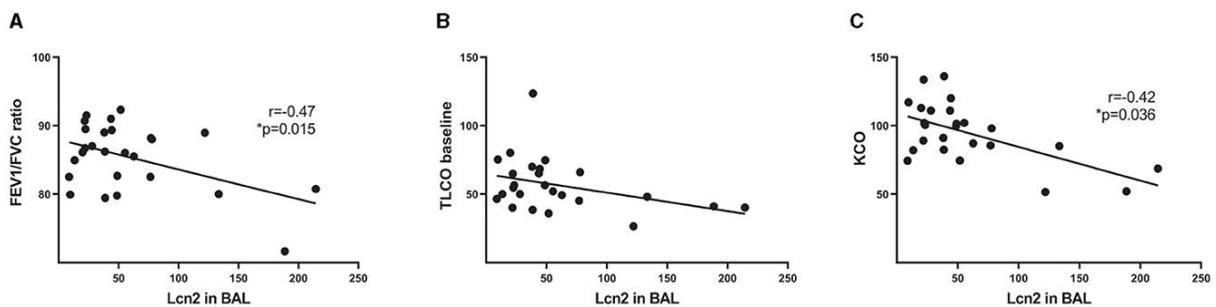


Figure 22 LCN2 BALF levels negatively relate to lung function metrics. A–C. Spearman’s correlation of ELISA-defined LCN2 levels with (A) forced expiratory volume (FEV) to forced vital capacity (FVC) ratio, (B) transfer capacity of the lung for carbon monoxide uptake (TLCO) and (C) carbon monoxide transfer coefficient (KCO). Statistical significance was delimited by absolute Spearman’s $r > 0.4$ and $*p < 0.05$. Adopted from (Galaris, Fanidis et al. 2023).

Table 14 Characteristics of the IPF patients cohort. Adopted from (Galaris, Fanidis et al. 2023).

Characteristic	IPF (n = 26)
Age (yr) (Mean ± SD)	72.8 ± 7.3
Gender, n (%)	
Male	25 (96.1%)
Female	1 (3.9%)
Pulmonary function tests (mean ± SD)	
DLCO%	56.2 ± 19.4
FEV1/FVC%	85.4 ± 4.7
KCO	94.1 ± 21.5
Hematological analysis (%) (mean ± SD)	
Macrophages	83.1 ± 9.8
Lymphocytes	7.6 ± 7.6*
Polymorphonuclear	7.5 ± 6.4
Eosinophils	1.3 ± 1.61
BALF LCN2 (ng/mL) eosinophils	58.9 ± 52.3

*FVC%, Forced vital capacity percent predicted; DLCO%, Carbon monoxide diffusing capacity percent predicted; FEV1%, Forced expiratory volume in 1-s percent predicted.

3.6.2 Mouse *Lcn2* expression patterns

Samples of BLM-induced PF and their control counterparts were mined through Fibromine to explore *Lcn2* differential expression between phenotypes (Table S12). As in the case of human species, we found *Lcn2* over-expressed during the fibrotic phase of the model in most of the datasets (Figure 23A-B). Re-analysis of murine scRNA-seq data (Strunz, Simon et al. 2020) revealed that *Lcn2* is produced primarily by epithelial cells and neutrophils (Figure 23C-D), while CellMarker2.0 database query (Hu, Li et al. 2023) proposed *Lcn2* as a marker of both neutrophils and AT2 cells (Table S14). More crucially, scRNA-seq data suggested that *Lcn2* mRNA quantities peak during the initial model stages that are characterized by acute inflammation, neutrophil accumulation and epithelial damage (Figure 23E-G).

In order to replicate the aforementioned findings, we administered BLM via oropharyngeal aspiration to 8-10 weeks old C57Bl6 mice, sacrificed at three time-points corresponding to the inflammation and fibrotic disease stages. Post to multiple model validation experiments (please refer to respective publication), we examined *Lcn2* expression in lung tissue via Q-RT-PCR and validated its detection in all disease phases (Figure 24K). Although BAL fluid *Lcn2* concentrations followed a similar motif (Figure 24L), serum protein levels differed from those of control samples only during acute inflammation (day 3) (Figure 24M), a phase rich in neutrophils as suggested by subsequent FACS analysis (Figure 24N). Following the acute inflammation, Western blot analysis revealed that *Lcn2* is detectable on day 14 post-BLM administration (Figure 24P-Q).

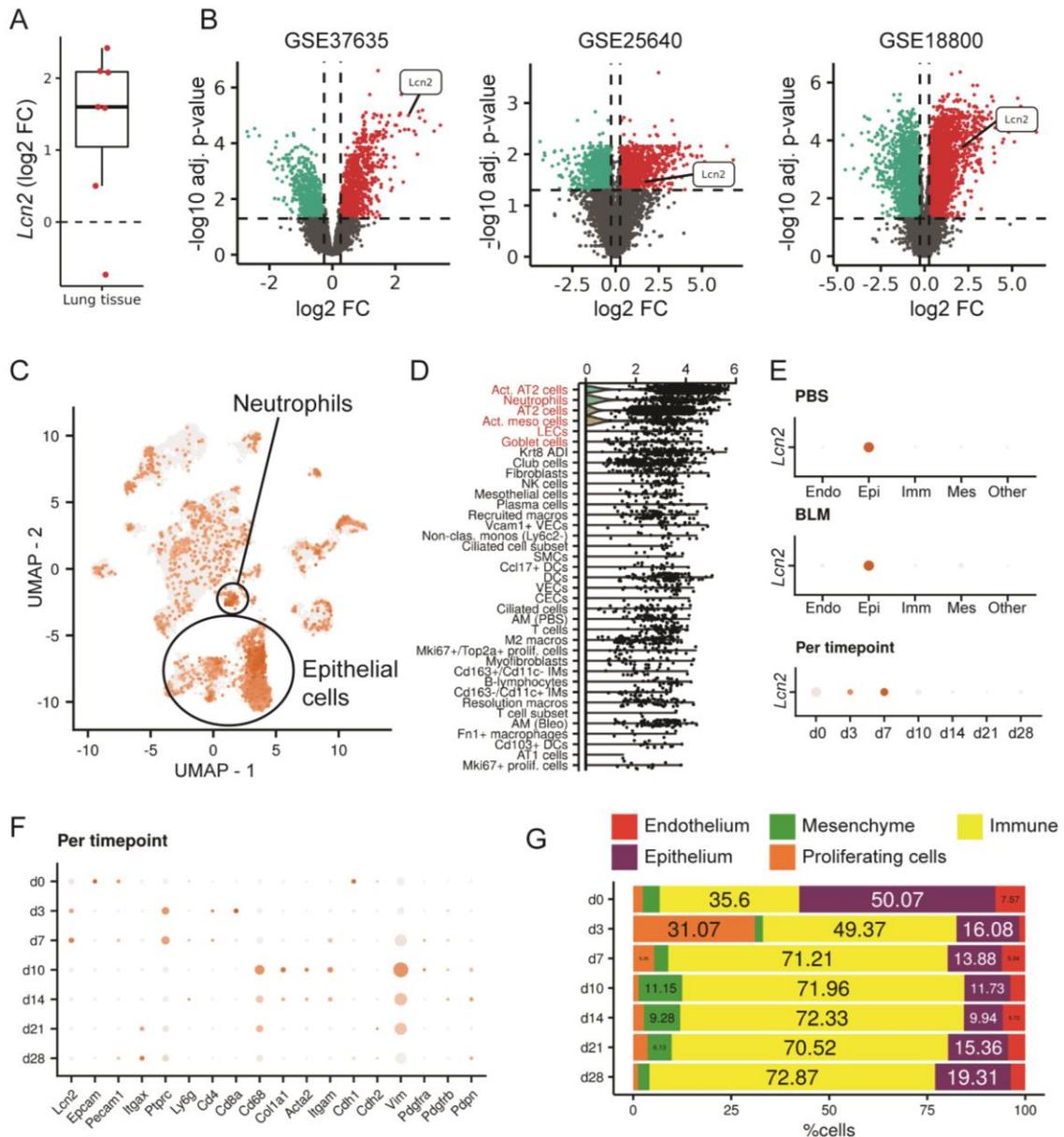


Figure 23 *Lcn2* expression pattern in mouse fibrosis model. **A.** *Lcn2* differential expression in several transcriptomic datasets comparing BLM- and non-treated mice (Table S12). **B.** Volcano plots of (A)-included datasets. **C.** *Lcn2* expression in the mouse lung at single cell resolution. **D.** *Lcn2* expression pattern in cell types ordered by decreasing level of importance. Transcription comparison demonstrated *Lcn2* as a marker gene of the red colored cell groups (Wilcoxon rank-sum test; FC > 1.2; Bonferroni adjusted p < 0.05). **E.** *Lcn2* epithelial origin can be supported through separate examination of control (PBS) and fibrotic (BLM) cells. **F.** Time-point-specific expression of cell population markers. **G.** Changes of mouse lung major cell populations in respect to BLM administration as described by scRNA-seq data. Adopted from (Galaris, Fanidis et al. 2023).

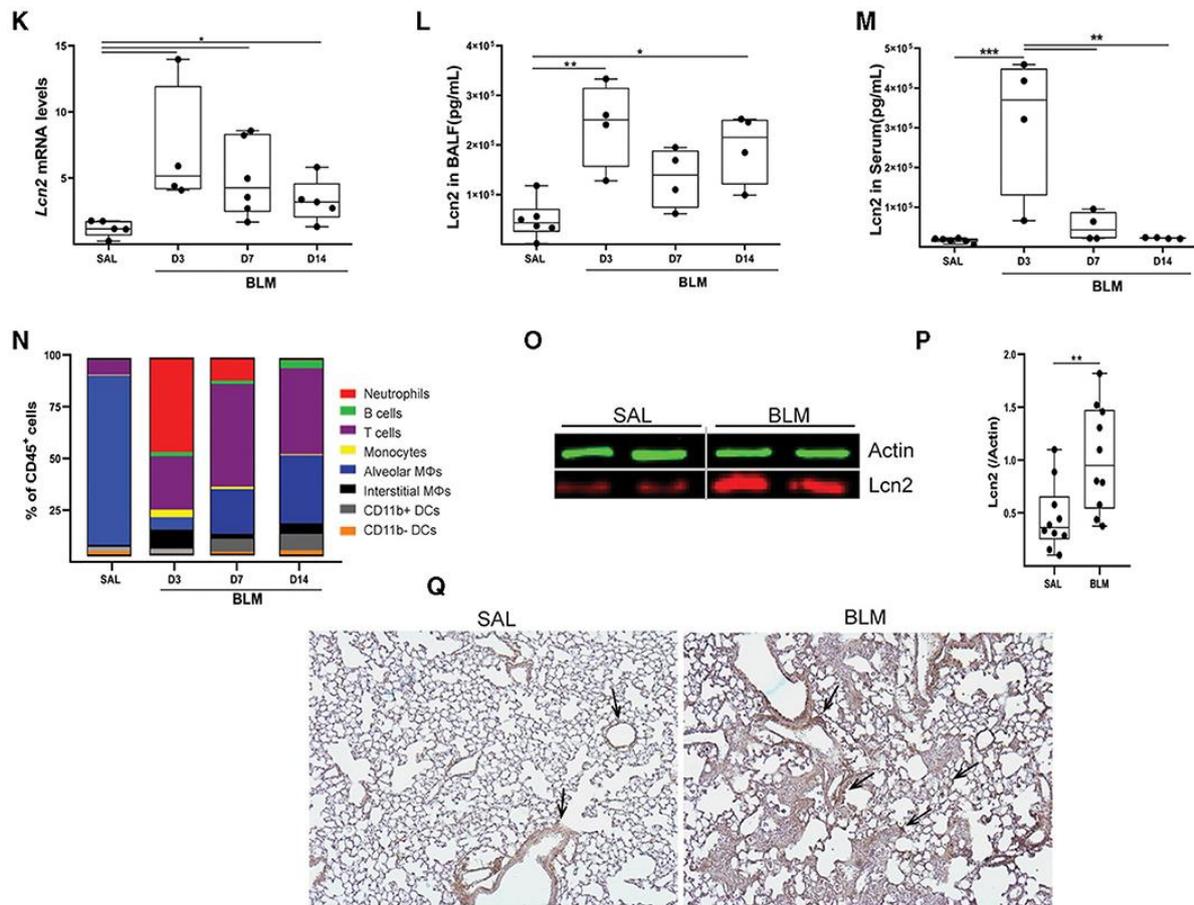


Figure 24 *Lcn2* levels are increased in the lungs of the BLM mouse model. **K.** Q-RT-PCR measured *Lcn2* expression presented as fold change over control; expression was normalized over that of *B2m*. **L-M.** ELISA-measured *Lcn2* quantities in BAL fluid (L) and serum (M) of mice at 3, 7, and 14 d post-BLM aspiration (one-way ANOVA, */**/** denote $p < 0.05/0.01/0.001$ respectively). **N.** Composition of the immune compartment in the murine lung post-BLM administration. **O.** Representative Western blot of *Lcn2* quantities 14 d post-BLM. **P.** Densitometry analysis of *Lcn2*, normalized to Actin expression (green); cumulative result from two independent experiments (unpaired t-test; ** $p < 0.01$). **Q.** Representative images of two independent immunohistochemistries for *Lcn2* in saline (SAL) and BLM treated murine lung tissue ($\times 10$). Adapted from (Galaris, Fanidis et al. 2023).

In an attempt to support *Lcn2* as an indicator of pulmonary inflammation, we measured *Lcn2* levels during an LPS-induced acute lung injury (ALI). ALI establishment and inflammatory cells infiltration (Figure 25A-B, E) were related to increased *Lcn2* transcription and translation (Figure 25C-D). This relationship was mirrored in both BAL fluid and sera samples of the same animals (Figure 25F-G).

In conclusion, both *in silico* and *ex vivo* analyses support that *Lcn2* is a marker of acute lung inflammation occurring either independently or as the first stage of PF.

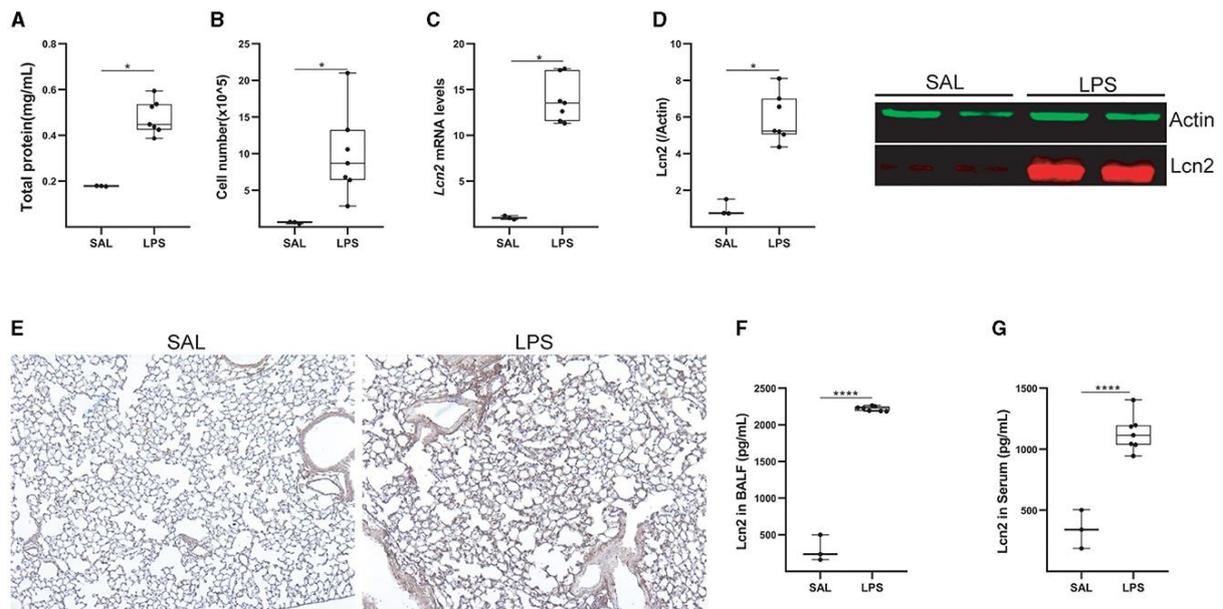


Figure 25 *Lcn2* levels are increased during acute lung injury episodes. **A.** BAL fluid total protein concentration (Bradford assay). **B.** Difference in BAL fluid inflammatory cell numbers from saline and LPS-treated mice (hemacytometer). **C.** Q-RT-PCR measured *Lcn2* expression normalized over that of *B2m* and iconized as fold change over control (representative results from three independent experiments). **D.** Western blot of *Lcn2* in mouse lungs suffering from LPS-induced ALI, accompanied by Actin-normalized densitometry analysis. **E.** Representative immunohistochemistry images for *Lcn2* in SAL and LPS-treated mouse lungs ($\times 10$). **F-G.** ELISA-measured BAL fluid and serum *Lcn2* levels (Mann–Whitney test; */**** denote $p < 0.05/0.0001$). Adopted from (Galaris, Fanidis et al. 2023).

3.6.3 Genetic approach of *Lcn2* role in murine lung pathology

Subsequently, we genetically knocked out *Lcn2* in order to dissect its molecular function. Despite the successful knock-out (Figure 26A-B), we observed no effect on several BLM-induced changes, such as weight loss (Figure 26C) and the aberrant production of soluble collagen (Figure 26F). In addition, lung structure was not altered (Figure 26G-H). Nevertheless, BLM inability to cause significant changes in lung function of *Lcn2* knock-out mice (Figure 26I-J) suggests that as in humans, increased levels of *Lcn2* negatively correlate with proper lung function. As *Lcn2* has been previously connected to obesity (Jaberi, Cohen et al. 2021) which in turn has been related to IPF (Guo, Sunil et al. 2022), we then examined *Lcn2* contribution in IPF establishment in obese (HFD-fed) compared to lean (control diet-fed) mice. In accordance to the PF model, no pathology-related changes were observed (Figure S5A-D). Last, motivated by the aforementioned increase in *Lcn2* expression during acute inflammation (Figure 24-25), we treated *Lcn2*^{-/-} mice with LPS to observe feature's effect on an active inflammation. Although knock-out animals suffered from increased pulmonary edema (Figure 27A), no changes in inflammation were detected (Figure 27B-C).

To conclude with, although initial analyses of both human and murine samples do pinpoint an increase in *LCN2/Lcn2* expression during acute inflammation and fibrosis, gene knock-out does not seem to have any significant effect on neither BLM-induced PF nor LPS-induced inflammation. Further experimentations are necessary to interpret *LCN2* up-regulation during divergence from steady state.

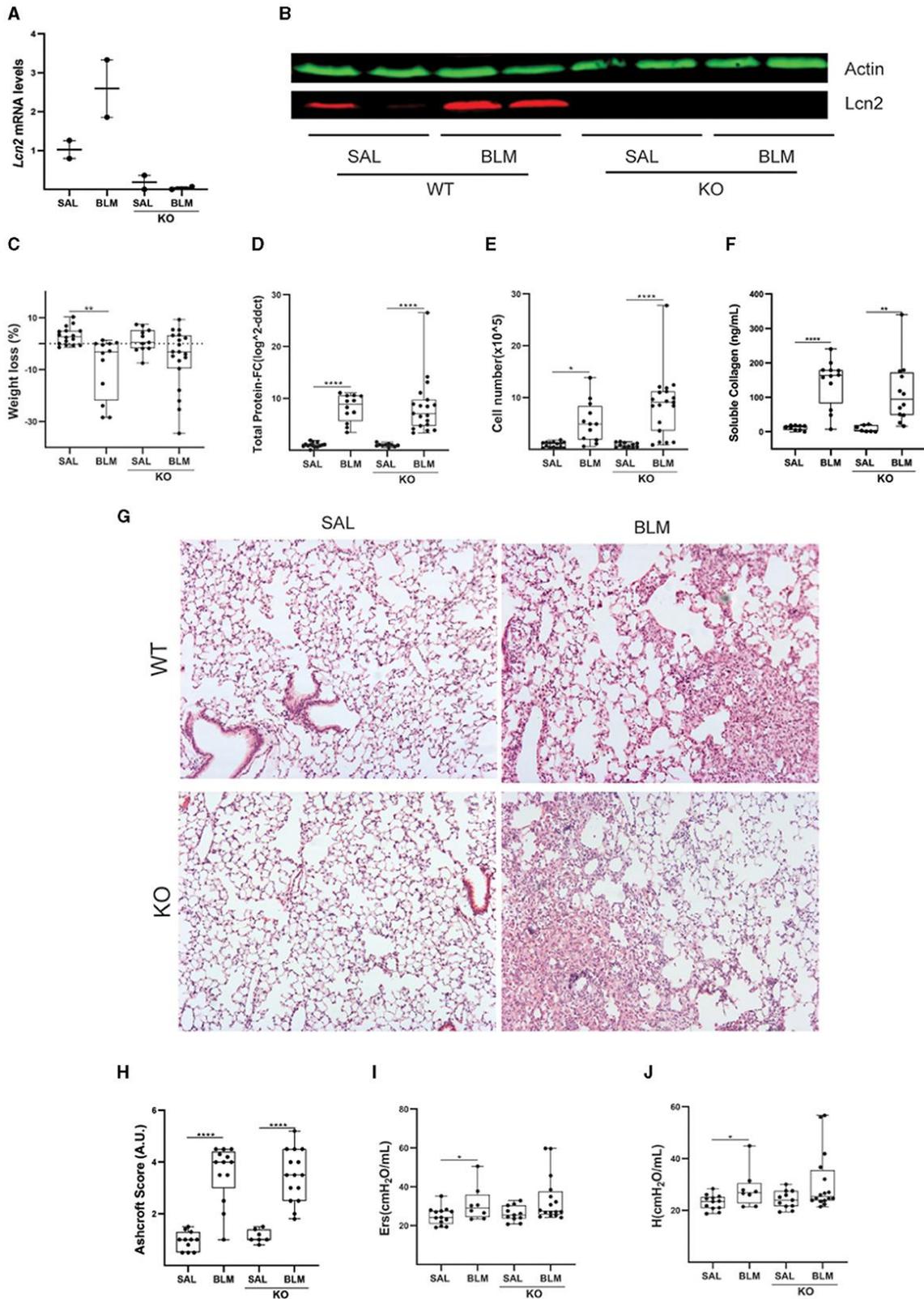


Figure 26 BLM-model is only slightly affected by *Lcn2* genetic deletion. **A.** Q-RT-PCR-examined *Lcn2* expression normalized over *B2m* expression and iconized as fold change over control. **B.** Representative Western blot of *Lcn2* in WT and KO lungs from BLM-treated mice proves global *Lcn2* depletion in KO mice. **C.** Weight loss post-BLM treatment. **D.** BAL fluid total protein (Bradford assay). **E.** Number of BAL fluid inflammatory cells (hemacytometer). **F.** Levels of BAL fluid soluble collagen detected via Direct Red assay. **G.** Representative H&E-stained lung sections ($\times 10$). **H.** Disease severity

as reflected by Ashcroft scoring. **I-J.** FlexiVent-measured metrics of respiratory function (one-way ANOVA; */**/**** denotes $p < 0.05/0.01/0.0001$). Adopted from (Galaris, Fanidis et al. 2023).

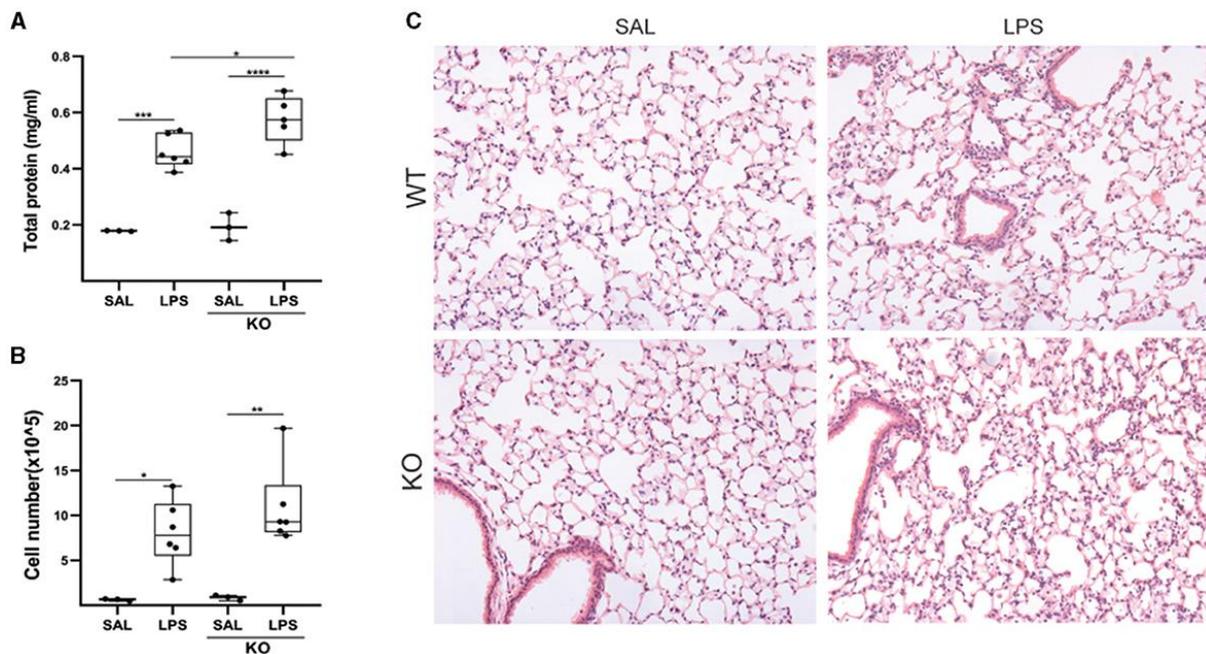


Figure 27 LPS-triggered pulmonary inflammation was marginally affected by *Lcn2* genetic deletion. **A.** BAL fluid total protein (Bradford assay). **B.** Number of BAL fluid inflammatory cells (hemacytometer). **C.** Representative H&E-stained lung sections from of WT and *Lcn2* KO animals ($\times 10$) (one-way ANOVA; */**/****/**** denotes $p < 0.05/0.01/0.001/0.0001$). Adopted from (Galaris, Fanidis et al. 2023).

3.7 TKS5 and fibroblast migration

TKS5 is a structural protein of podosomes, necessary accessories for cells migration (Introduction). Increased TKS5 expression has been reported in various cancer types (Murphy and Courtneidge 2011, Hoshino, Branch et al. 2013, Paterson and Courtneidge 2018), while it was also suggested to mediate lung tumor metastasis (Li, Chen et al. 2013). PF is among the higher risk factors for lung cancer development and ECM invasion is just one of the multiple commonalities shared between lung fibroblasts and cancer cells (Karampitsakos, Tzilas et al. 2017). Motivated by the above observations, we decided to examine the involvement of TKS5 in PF.

3.7.1 *TKS5* expression is upregulated during PF

Initially, we computationally assessed *TKS5* expression in human lungs during both fibrosis and steady state. Using Fibromine (Fanidis, Moulos et al. 2021), we observed a *TKS5* over-expression in almost all IPF_vs_Control bulk RNA-seq lung datasets (Figure 26A; Table S15) incorporating three of the biggest ones (Figure 28B; Figure S6A,C). In addition, *TKS5* transcription positively correlated with that of *COL1A1* (Figure 28C; Figure S6B,D), strengthening the connection between *TKS5* deregulation and PF establishment.

To support our *in silico* findings we applied Q-RT-PCR which validated the increased *TKS5* mRNA levels in IPF lung tissue ($n=20$) compared to both control ($n=9$) and COPD-suffering individuals ($n=19$) (Figure 28D; Table S16). Next, by immunostaining lung sections we identified TKS5 in the alveolar epithelium and the fibrotic areas of IPF/UIP ($n=3$) patients, but

not in control samples (n=3) (Figure 28E). Re-analysis of a scRNA-seq dataset (Reyffman, Walter et al. 2018) located strong *TKS5* expression in epithelial cell types, basal cells and primarily fibroblasts (Figure S6E-F), with *TKS5*⁺ fibroblasts belonging to a *COL1A1*⁺ subpopulation (Figure S6G-F).

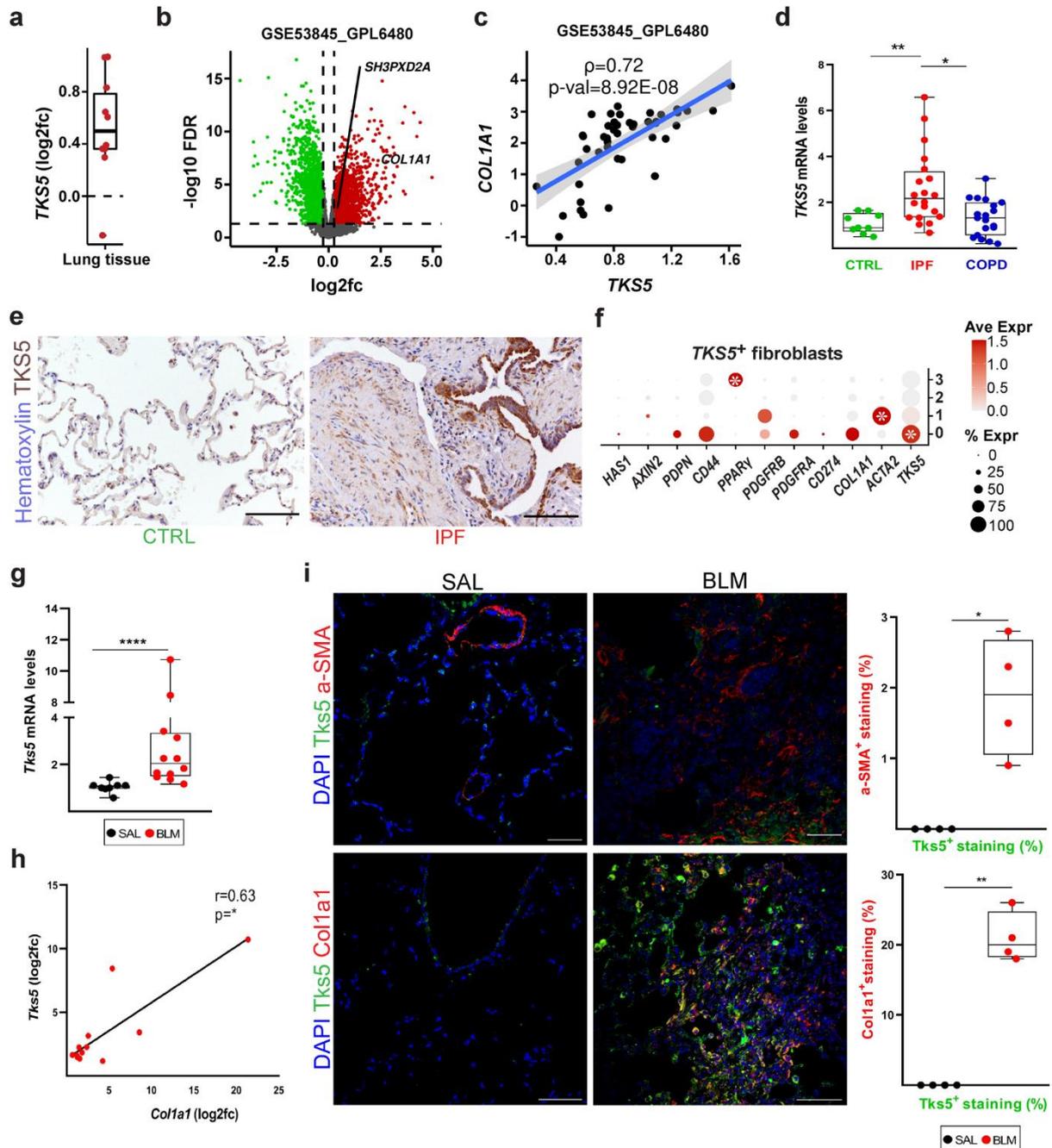


Figure 28 *TKS5* expression is increased during pulmonary fibrosis. **A.** *TKS5* expression in lung tissue of IPF patients compared (log₂FC) to that of controls individuals as represented in nine Fibromine-hosted datasets (Table S15). **B.** Volcano plot from a large dataset of (A) (FC > 1.2, FDR < 0.05). **C.** *TKS5* and *COL1A1* expression correlation in the dataset of (B) with a fitted linear model and 95% CI (two-tailed Spearman's test; $\rho > 0.6$; $p = 8.92\text{E-}08$). **D.** Q-RT-PCR revealed *TKS5* over-expression in IPF/UIP lung tissue samples (n = 20) compared to COPD (n = 19) and tumor-free control lung tissue (n = 9) ($r_2 = 0.98$, $E = 97\%$) (Table S16) (normalization to the expression of *B2M*; presentation as fold change to control values). Statistical significance was established with two-tailed

Kruskal-Wallis test (**p = 0.0076, *p = 0.0129). **E.** Representative immunostaining images demonstrating a TKS5 increase in fibrotic lungs (n = 3; scale bars = 50µm). **F.** COL1A1+ lung fibroblasts are the main producers of TKS5 as witnessed from the analysis of a scRNA-seq dataset (Reyfman, Walter et al. 2018) (Wilcoxon Rank Sum test; *FC > 1.2; Bonferroni corrected p = 8.9E-12 / 1.1E-10 / 2.1E-3 from left to right). **G-H.** Q-RT-PCR-measured expression of *Tks5* and *Colla1* normalized over that of *B2m* and iconized as fold change (log2) over control (n = 8/12) (r2 = 0.89/0.93; E = 103%/96%); cumulative result from 3 experiments (two-tailed Mann Witney test; ****p < 0.0001). **H.** Correlation of *Tks5* and *Colla1* mRNA levels in the same samples (two-tailed Spearman's test; *p = 0.0323; r = 0.63). **I.** Representative photographs of double immunostaining for Tks5 and aSMA (Acta2) or Colla1, followed by their respective Image J derived quantification (n = 4); scale bars=50 µm (two-tailed Welch's test; *p = 0.0211, **p = 0.0013). Adopted from (Barbayianni, Kanellopoulou et al. 2023).

Subsequently, we set to generalize our findings and assess the eligibility of the BLM mouse model for *TKS5* exploration. Towards that purpose we examined publicly available murine model datasets. As in the case of the human disease, *Tks5* mRNA levels increased in BLM-treated compared to untreated animals, with this up-regulation correlating with *Colla1* expression (Figure 28G-H). Immunostaining ascribed Tks5 expression in the alveolar epithelium and lung fibrotic areas (Figure 28I) supporting our findings in human samples. Last, Colla1, aSMA and Tks5 stainings comparison revealed that 1/5 of the Colla1+ fibroblasts were also Tks5+ in contrast to only 2% of their aSMA+ counterparts (Figure 28I).

Collectively, fibrosis affected human and mouse lung tissue over-expresses *TKS5/Tks5* similarly to *COL1A1/Colla1*. *TKS5/Tks5* transcription can be located to certain pulmonary fibroblast subtypes.

3.7.2 Lung fibroblasts intrinsically create TGFβ-induced TKS5 podosome rosettes

Based on the above findings and the already known pro-fibrotic effects of TGFβ on pulmonary fibroblasts (Ye and Hu 2021), we next evaluated whether TGFβ signaling does affect TKS5 levels and function. To begin with, TGFβ treatment of primary normal human lung fibroblasts (NHLFs) not only induced *TKS5* transcription (Figure 29A), but even echoed the *TKS5-COL1A1* relationship seen during active fibrosis (Figure 29B). The same observations were made in various other experimental contexts (Figure S7A-D). In addition, NHLF *in vitro* treatment with TGFβ induced podosomes formation (Figure 29C-F). MMP9 detected at the same structures (Figure 29G-H) potentially participates to the increased chemical degradation of fluorescein-conjugated gelatin substrate (Figure 29I-J), further supporting the structures' identity as that of functioning podosomes.

Next we set to examine whether the IPF lung pro-fibrotic milieu can *in vivo* induce the formation of podosomes. Towards that purpose, we cultured IPF-originating lung fibroblasts

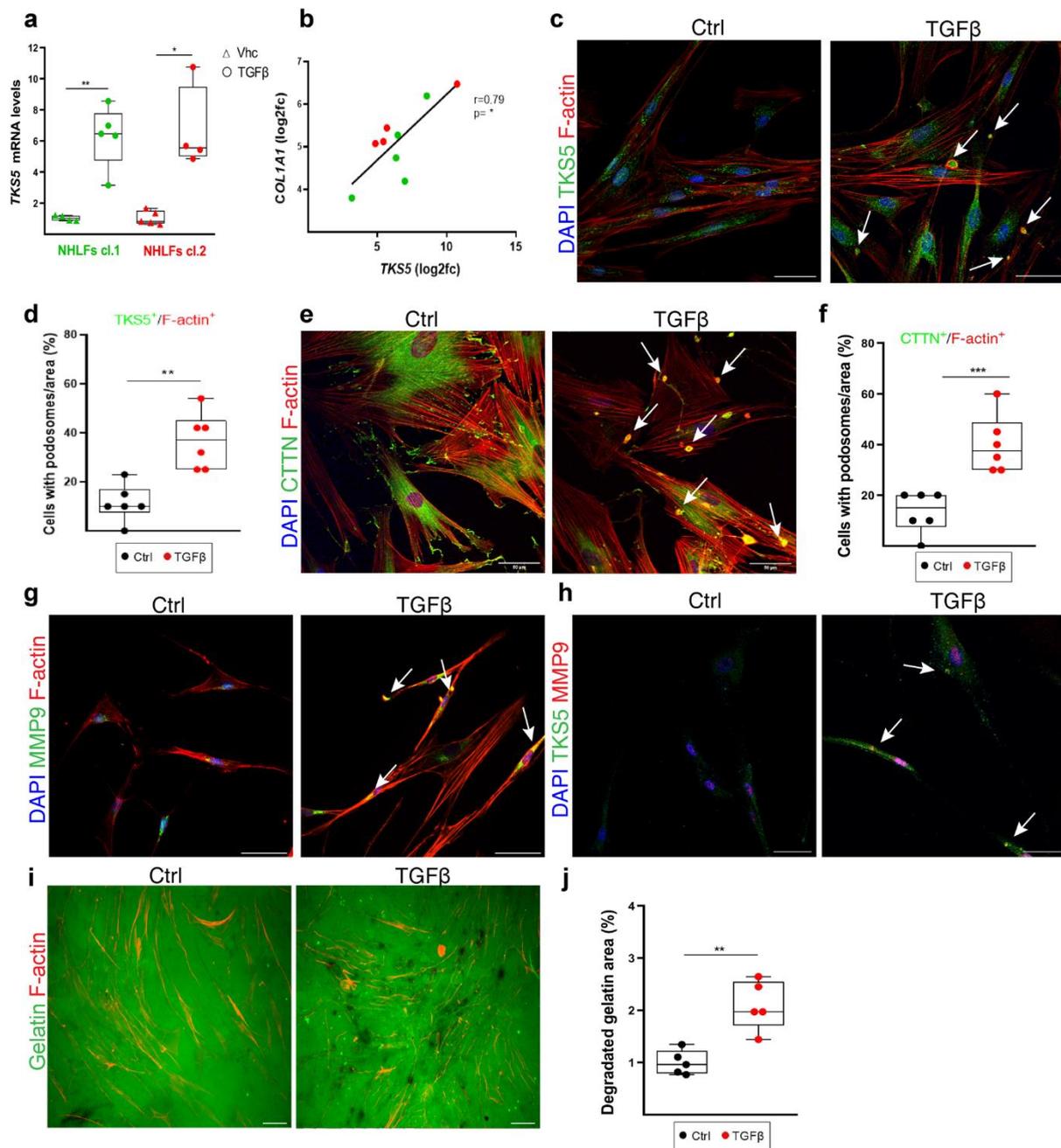


Figure 29 TGF β stimulates normal human lung fibroblasts to create podosomes. **A-B.** Q-RT-PCR measured expression of *TKS5* and *COL1A1* ($r_2 = 0,94/0,92$; $E = 98,3\%/93\%$ respectively) in two NHLF clones (cl.1, cl.2) normalized to *B2M* expression and presented as fold change over control; $n = 4/5/4/4$; (two-tailed Welch's test (a/cl.1) and two-tailed Mann Whitney test (a/l.2); $**p = 0.0012$, $*p = 0.0159$). **B.** Two-tailed Pearson's correlation of *TKS5* and *COL1A1* expression values in the same samples ($*p = 0.0116$). **C-J.** Representative composite images from double immunostaining, and respective quantifications, for: (C) F-actin/*TKS5*, (E) F-actin/*Cortactin (CTTN)*, (G) F-actin/*MMP9* and (H) *TKS5*/*MMP9*. Scale bars 50 μm ; arrows point to representative podosomes. (D, F) Quantification of podosome-containing cells per optical field ($n = 6$) (two-tailed t-test; $**p = 0.0011$, $***p = 0.0009$). **I.** Representative TGF β -induced degradation (black holes) of a fluorescein-conjugated gelatin substrate. **J.** Quantification of gelatin degradation area, normalized to control ($n = 5$) (two-tailed t-test; $**p = 0.0016$). Adopted from (Barbayianni, Kanellopoulou et al. 2023).

(HLFs) (Table S17) without any treatment and stained for podosomes, while various NHLF cell lines cultured under the same conditions were used as control. Podosome rosettes were identified in the pathologic HLFs irrespective of their culture density (Figure 30A-D). Interestingly, podosomes were structurally the same with the TGF β -induced ones and they also persisted for a long period in *ex vivo* cultures. Similarly yet more effectively to TGF β -stimulated NHLFs, IPF HLFs lysed a fluorescein-conjugated gelatin substrate that we used (Figure 30E-F).

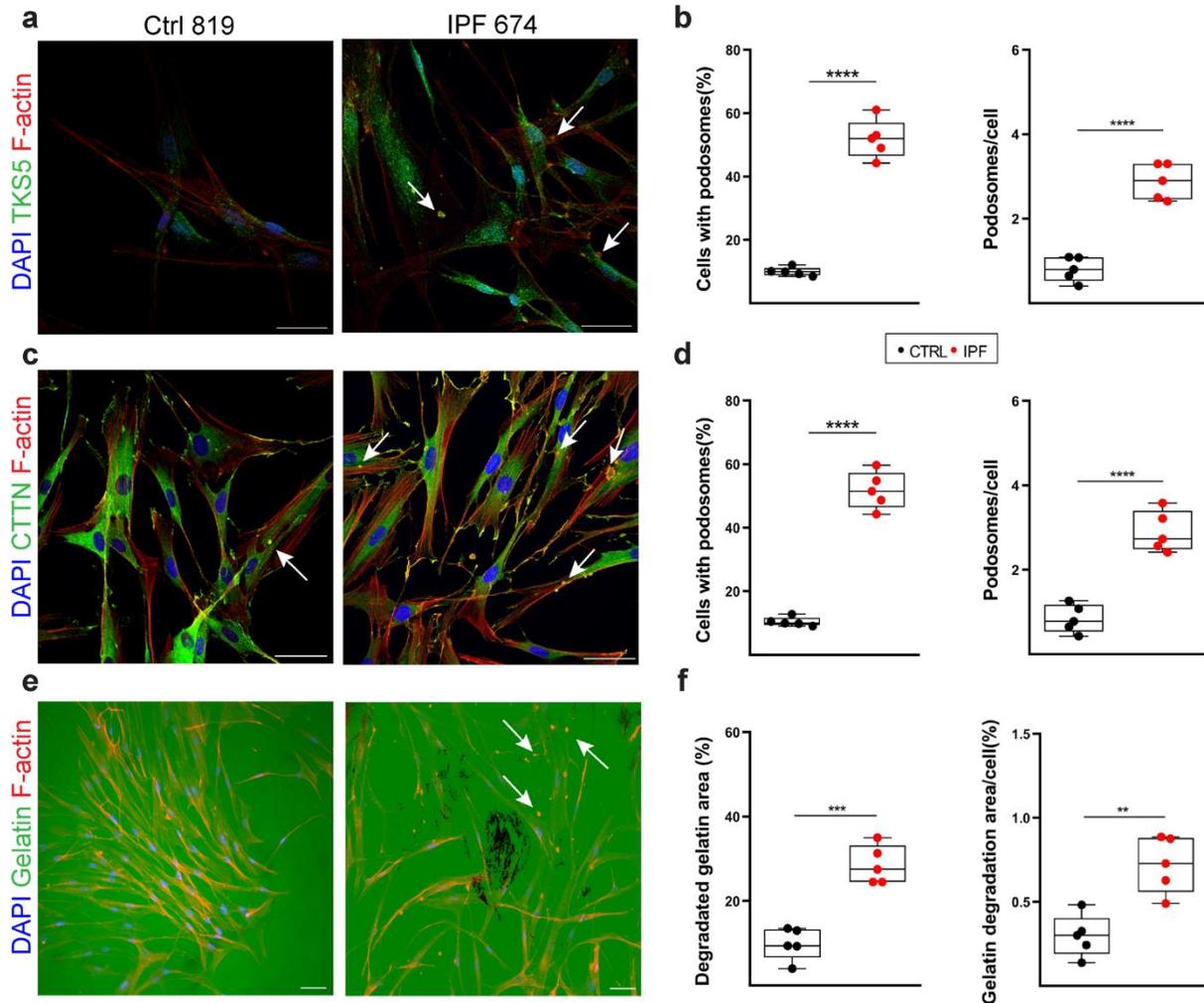


Figure 30 Human lung fibroblasts homeostatically produce podosomes. **A, C.** Serum starved, sub-confluent primary IPF-HLFs and NHLFs immunostained for F-actin and (A) TKS5 or (C) cortactin. Counter stained with DAPI; n = 5; scale bars = 50 μ m. **B, D.** Cumulative number of podosome-containing cells (%) and podosomes number per cell per optical field, respectively (two-tailed t-test (b) or two-tailed Welch's test (d); ****p < 0.0001). **E.** Staining of the same clones cultured on a fluorescein-conjugated gelatin substrate for F-actin. Counter staining with DAPI. **F.** Cumulative quantification the degraded gelatin (%) for all clones and quantification of gelatin degradation area per cell (ImageJ; two-tailed t-test; *** p = 0.0001/**p = 0.0020). Adopted from (Barbayianni, Kanellopoulou et al. 2023).

In support of the above findings, TGF β -treatment of primary normal mouse lung fibroblasts (NMLFs) stimulated *Tks5* transcription in a fashion similar to that of *Colla1* (Figure S8A-B). Moreover, we were able to identify podosome rosettes (Figure S8C-D) and lysis of a fluorescein-conjugated gelatin substrate (Figure S8E-F). Matching results were retrieved from

3T3 embryonic fibroblasts (Figure S8G-K). As in the case of their human homologues, primary lung fibroblasts isolated from BLM-treated animals were characterized by an increased transcription of *Tks5*, *Colla1* and *Mmp9* (Figure S8L-N), as well as by podosome rosettes formation (Figure S8O-P).

The aforementioned results suggest that both TGF β treatment and the IPF/BLM pro-fibrotic milieu stimulate lung fibroblasts to up-regulate the expression of *TKS5/Tks5* and the development of likely imprinted podosome rosettes.

3.7.3 *Tks5* haploinsufficiency diminish BLM-induced lung fibrosis

Subsequently, we set to evaluate the effects of *Tks5* during PF. Working towards that goal we created a C57Bl6/J *Tks5*^{+/-} mouse model which we treated with BLM (Figure 31A-B). WT littermates were used as control. For details regarding knockout mice creation, please, refer to (Barbayianni, Kanellopoulou et al. 2023). Animals were sacrificed 14 days after BLM administration, at the peak of the modeled fibrosis (Barbayianni, Ninou et al. 2018). From some initial analyses, *Tks5*^{+/-} mice did not lose any weight in contrast to WT ones (Figure 31C), while vascular leak and pulmonary oedema were definitively reduced in the former (Figure 31D). Moreover, inflammatory cells, BAL fluid soluble collagen levels and *Colla1* expression were significantly reduced in the knockout animals (Figure 31E-H). Histologically, we discovered less collagen being deposited in BLM-treated knockout mice (Figure 31I), in parallel to less fibrotic regions both in the parenchyma and peribronchial areas (Figure 31J). Observations from *ex vivo* cultures of Precision Cut Lung Slices (PCLS) from the same mice were in line with the above findings (Figure 31I). Last, lung function was not significantly affected by BLM in *Tks5*^{+/-} animals compared to their WT littermates (Figure 31K-M).

Hence, *Tks5* expression and by extension podosomes seem to actively participate in BLM-induced PF, with these observations possibly generalizable to the human IPF lung.

3.7.4 *Tks5* haploinsufficiency diminish BLM-induced lung fibrosis

With the aim of functionally examining the anti-fibrotic properties of *Tks5* haploinsufficient mice, we collected primary lung fibroblasts from littermate WT and *Tks5*^{+/-} mice and treated them with TGF β . *Tks5*^{+/-} cells (Figure 32A) had a smaller number of podosomes (Figure 32B-C) and decreased proliferation rates (Figure 32D). To evaluate podosomes' ability to actively support cells' movement through ECM, we cultured murine lung fibroblasts on murine lung-originating acellular ECM (Figure S9A-B) in a transwell invasion chamber for 6h (Figure 32E). *Tks5*^{+/-} originating cells were characterized by a decreased number of TGF β -induced podosome rosettes, a reduction which was accompanied by a decline in cells' invasion potential (Figure 32F). Last, BLM-treated *Tks5*^{+/-} fibroblasts had less podosomes compared to their WT counterparts (Figure 32G-H) in parallel with a reduced capacity for acellular ECM invasion (Figure 32I).

In search of a mechanism of action, we once again treated murine WT and *Tks5*^{+/-} lung fibroblasts with TGF β and used 3' UTR RNAseq (Quant-Seq LEXOGEN) to interrogate their transcriptional profile. DEA identified 3648 deregulated genes (FC>1.2, FDR corrected p<0.05) (Figure S10A), with 418 of them already known to be implicated in fibrosis as revealed

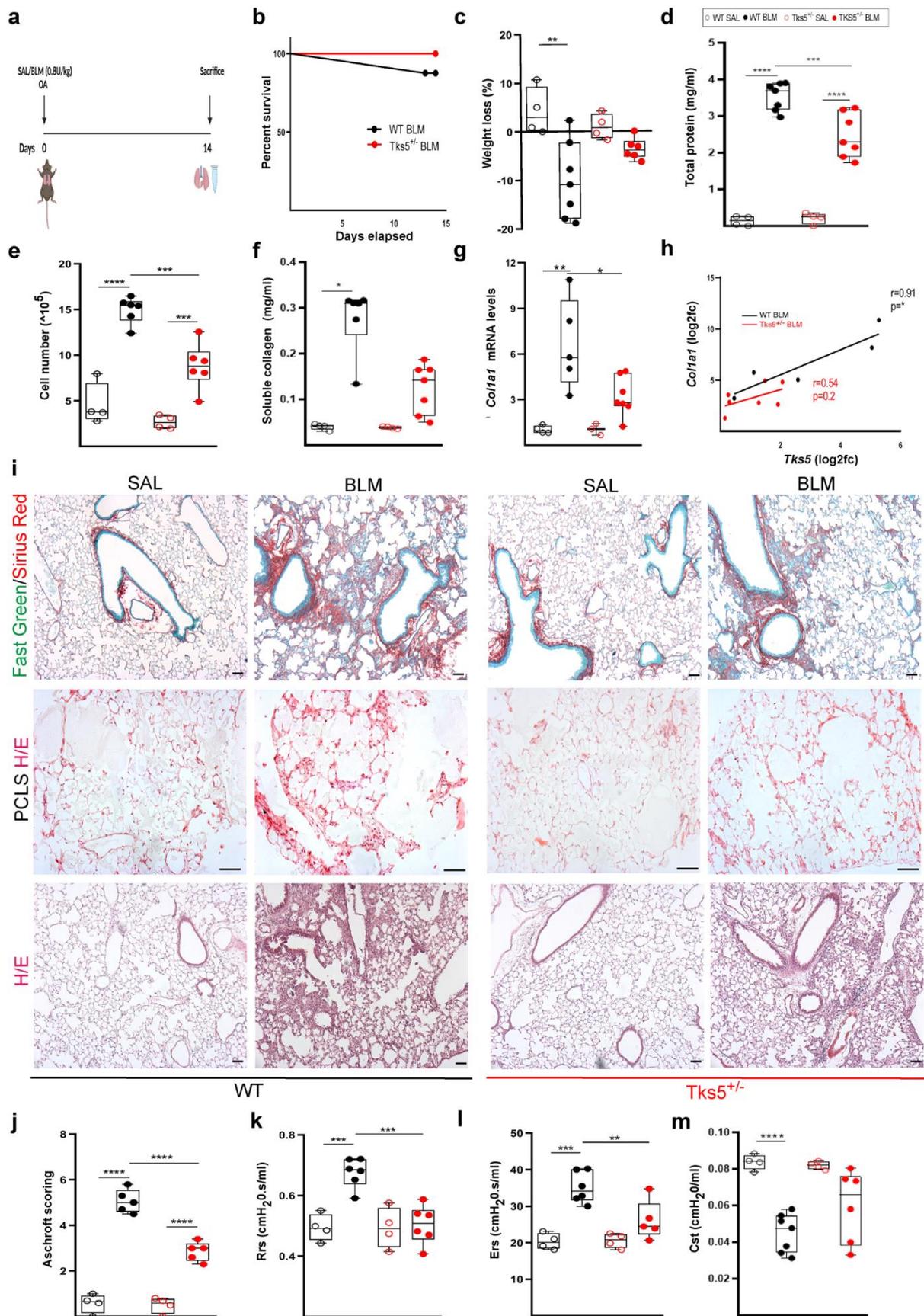


Figure 31 BLM-established pulmonary fibrosis is weakened in *Tks5* haploinsufficient animals. **A.** BLM-induced PF model protocol. **B.** Kaplan Meyer survival curve for wild type and *Tks5*

haploinsufficient animals post to BLM treatment. **C.** Changes in body weight after BLM administration (two-tailed one-way ANOVA; $**p = 0.031$). **D.** BAL fluid total protein concentration (Bradford assay; two-tailed one-way ANOVA; $****p < 0.0001$, $***p = 0.0009$). **E.** Number of inflammatory cells in BAL fluid (hemacytometer; two-tailed one-way ANOVA; $****p < 0.0001$, $***p = 0.0002/0.0008$). **F.** Levels of soluble collagen in BAL fluid (Direct red assay; two-tailed Kruskal Wallis; $*p = 0.0124$). **G-H.** Q-RT-PCR measured expression of *Tks5* and *Colla1* normalized over that of *B2m* and presented as fold change over control (two-tailed one-way ANOVA; $**p = 0.0012$, $*p = 0.0207$). **H.** Two-tailed Pearson's correlation of *Tks5* and *Colla1* expression in the same samples ($*p = 0.0342$; $r = 0.91/0.54$). **I.** Representative images from: murine lung sections stained with Fast Green/Sirius Red (first row), precision cut lung slices stained with Hematoxylin & Eosin (H&E) (second row) and lung sections stained with H&E (third row) (scale bars 50 μm). **J.** Fibrosis severity as defined by Ashcroft scoring H/E lung sections (two-tailed one-way ANOVA; $****p < 0.0001$). **K.** Mean respiratory system resistance (Rrs; two-tailed one-way ANOVA; $***p = 0.0008/0.0004$). **L.** Mean respiratory system elastance (Ers; two-tailed one-way ANOVA; $***p = 0.0002$, $**p = 0.0081$). **M.** Mean static lung compliance (Cst; two-tailed one-way ANOVA followed by Welch's correction; $****p < 0.0001$). Adopted from (Barbayianni, Kanellopoulou et al. 2023).

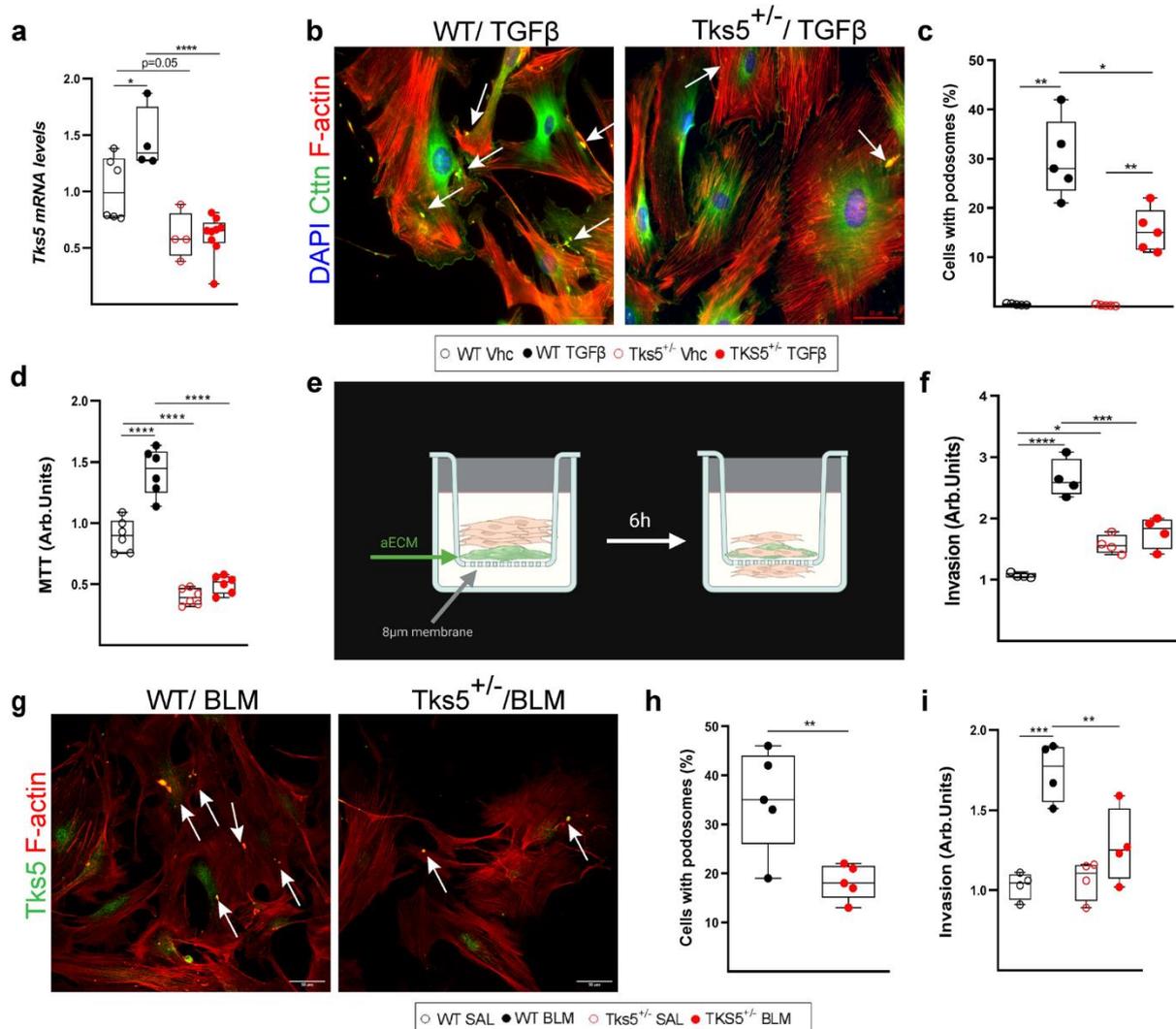


Figure 32 *Tks5* partial deletion attenuates fibroblasts ability to produce podosomes and invade the extracellular matrix. Serum-starved NMLFs from WT and *Tks5*^{+/-} mice treated with recombinant TGF- β 1 (10 ng/ml for 24 h). **A.** Q-RT-PCR measured *Tks5* expression normalized over that of *B2m* and presented as fold change over control (two-tailed one-way ANOVA; $*p = 0.0464$, $****p < 0.0001$). **B.**

Representative double immunostaining for F-actin and Cortactin counter-stained with DAPI; arrows indicate representative podosomes. **C.** Number of podosome containing cells per optical field (two-tailed one-way ANOVA followed by Welch's correction; * $p = 0.004/0.0053$ * $p = 0.0411$). **D.** TGF β -induced proliferation of NMLFs (MTT assay; two-tailed one-way ANOVA; **** $p < 0.0001$). **E.** Schematic of TGF β -stimulated lung fibroblasts invasion into acellular ECM (biorender.com). After 6h, cells located into the lower surface of the upper chamber were stained, lysed and their absorbance was measured. **F.** Invasion capacity of TGF- β treated NMLFs (transwell invasion assay; two-tailed one-way ANOVA; **** $p < 0.0001$, * $p = 0.0266$, *** $p = 0.0005$). **G.** Representative double immunostaining for F-actin and Tks5 in NMLFs retrieved from WT and Tks5 \pm mice, post BLM administration; arrows indicate representative podosomes (scale bars 50 μ m). **H.** Number of podosome-containing cells per optical field (two-tailed t-test; ** $p = 0.009$). **I.** Post-BLM invasion capacity of LFs (transwell invasion assay; two-tailed one-way ANOVA; *** $p = 0.0003$ ** $p = 0.0099$). Adopted from (Barbayianni, Kanellopoulou et al. 2023).

by text mining of PubMed-hosted publication abstracts. Examination of the DEG list for deregulated transcription factors revealed that *Stat1*, *Cebpa* and *Ar* were under-expressed in haploinsufficient cells along with multiple of their targets (Figure S10B). From a functional point-of-view, deregulated genes were enriched by ECM-related biological processes (Figure 33A), with "Collagen containing ECM" (GO:0062023) being the most prominent among them (Figure 33B, Figure S10C).

In extension of the above findings, BLM-treated Tks5 \pm lung fibroblasts, characterized by a smaller number of podosome formations and weakened acellular ECM (aECM) invasion capabilities (Figure 32G-I), had significantly reduced levels of Col1a1 (Figure 33C).

On the other hand, NMLFs grown on a Col1a1 $+$ BLM-treated mice originating aECM had stimulated *Tks5* expression and podosomes formation (Figure 33D-F). More interestingly, Col1a1 treatment increased Col1a1 production (Figure 33G), thus suggesting the existence of a signaling loop between ECM components and podosomes during lung fibroblasts activation.

To sum up, the above observations establish that Tks5 \pm animal models are at least partially protected from fibrosis due to impaired podosomes formation and subsequent reduction of ECM invasion. Importantly, both processes can be tracked down to the transcriptomic level thus revealing an impaired gene expression both regulated from and affecting lung ECM.

3.7.5 Inhibiting Src does reduce podosomes creation and diminish pulmonary fibrosis

Since the enhanced Tks5-dependent podosome formation is an inherent property of pro-fibrotic fibroblasts which if impaired can impede IPF progression, we wanted to discover drugs that could target this process. Towards that purpose, we have used CMap analysis (Lamb, Crawford et al. 2006) to compare the transcriptional profile of TGF β -treated Tks5 \pm lung fibroblasts with that of LINCS-L1000 database (Figure 34A). The latter resource provides thousands of gene expression profiles measured from multiple cell lines incubated to various small

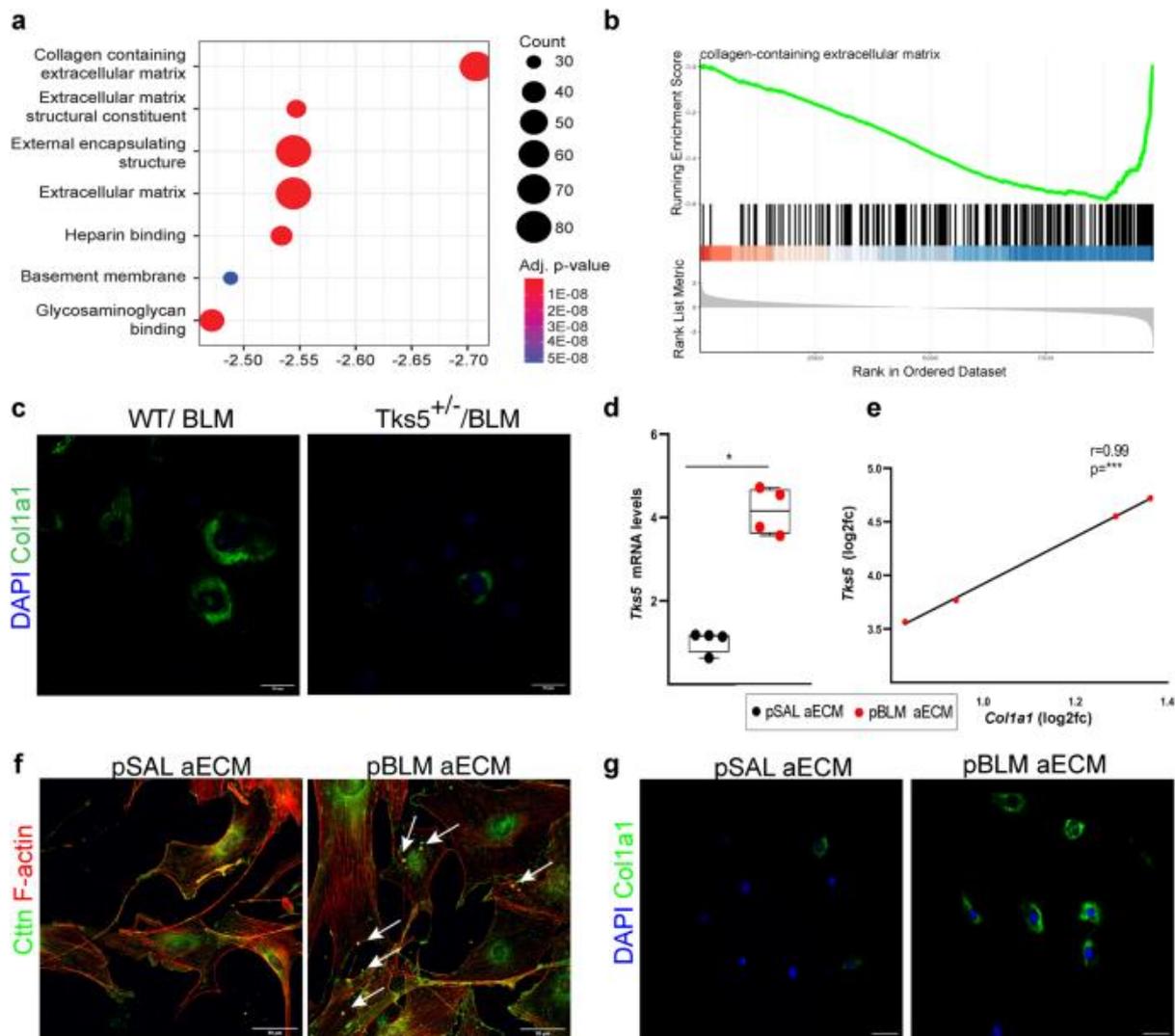


Figure 33 *Tks5* haploinsufficiency affects lung fibroblasts interaction with the extracellular matrix. **A.** ECM-related biological processes were enriched among down-regulated genes in TGF β treated *Tks5*^{+/-} lung fibroblasts compared to WT treated ones (fold change pre-ranked GSEA). **B.** The most enriched term among down-regulated genes. **C.** Serum-starved WT and *Tks5*^{+/-} originating lung fibroblasts immunostained for *Col1a1* and counter-stained with DAPI (scale bars = 50 μ m). **D-E.** Serum-starved WT primary lung fibroblasts grown in post saline (pSAL) and post bleomycin (pBLM) retrieved aECM. Q-RT-PCR-measured *Tks5* and *Col1a1* expression normalized over that of *B2m* and presented as fold change over control (two-tailed Mann-Whitney test; * $p = 0.0286$). **E.** Two-tailed Pearson's correlation between *Col1a1* and *Tks5* expression in the same samples (** $p = 0.0004$, $r = 0.99$). **F-G.** Representative double immunostaining for F-actin and Cortacin (Ctnn; F) or *Col1a1* (G) counter-stained with DAPI; arrows indicate representative podosomes; scale bars = 50 μ m. Adopted from (Barbayaanni, Kanellopoulou et al. 2023).

molecules and drugs (Subramanian, Narayan et al. 2017). During CMap analysis, the queried signature is juxtaposed to those in the database in search of a molecule that affects transcription in a similar or opposite fashion. In our case, we identified 15 signatures consistent to our knockout profile, several of which are known to have a positive impact on the health of animal disease models (Figure 34A; Table S18). Among the aforementioned signatures one can

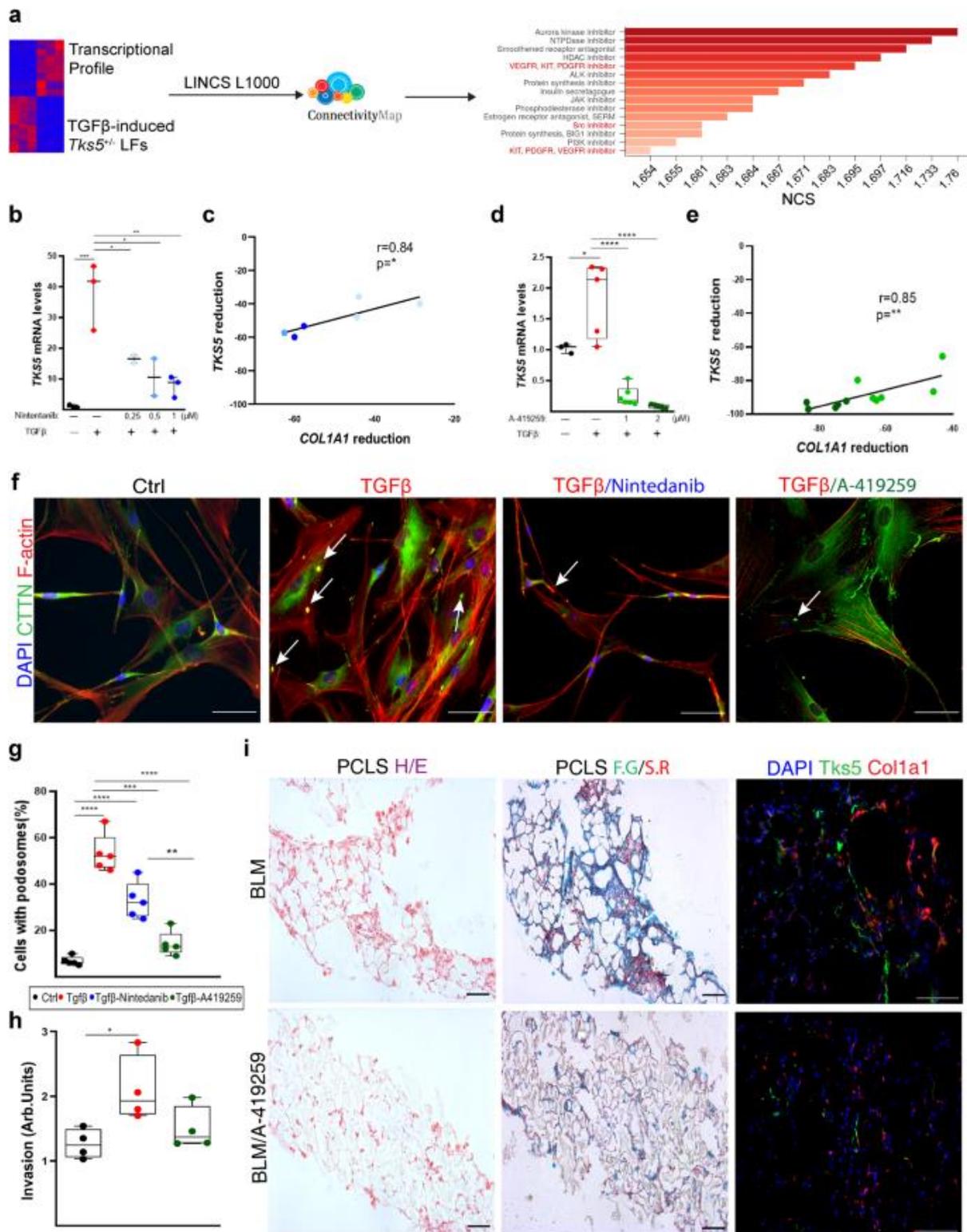


Figure 34 Src inhibition in human fibroblasts impairs podosome formation and extracellular matrix invading capabilities. Serum starved NHLFs pretreated for 1 h with A-419259 and Nintedanib prior to 24h treatment with recombinant human TGF β . **A.** CMap analysis pipeline and results. **B-C.** Q-RT-PCR measured expression of *TKS5* and *COL1A1* normalized to that of *B2M* and presented as fold change over control (two-tailed one-way ANOVA; *** $p = 0.0008$, * $p = 0.0397/0.0107$ ** $p = 0.003$). **C.** Correlation of *TKS5* and *COL1A1* expression in the same samples (two-tailed Pearson correlation; * $p = 0.0202$, $r = 0.84$). **D-E.** Q-RT-PCR measured expression of *TKS5* and *COL1A1* normalized to that of *B2M* and presented as fold change over control (two-tailed one-way ANOVA; * $p = 0.0202$, **** $p = 0.0001$).

< 0.0001). **E.** Correlation of *TKS5* and *COL1A1* expression in the same samples (two-tailed Pearson correlation; **p = 0.0018, r = 0.85). **F.** Representative double immunostaining for F-actin and Cortactin (cttn) counter stained with DAPI; arrows indicate representative podosomes; scale bars = 50 μ m. **G.** Quantification of podosome-containing cells per optical field (two-tailed one-way ANOVA; **** p < 0.0001, ***p = 0.0006, **p = 0.0014). **H.** Transwell invasion assay of NHLFs pre-treated with A-419259 and stimulated with TGF- β (two-tailed one-way ANOVA; *p = 0.0294). **I.** Pulmonary fibrosis reduction by Src-inhibition in mouse precision cut lung slices obtained post-BLM treatment (d11). A-419259 was administered in the first 24h after slicing for 3 consecutive days. PCLS stained with H&E, Fast green/Sirius red and double immunostaining for *Tks5* and *Colla1*; scale bars=50 μ m. Adopted from (Barbayianni, Kanellopoulou et al. 2023).

distinguish those of PDGF and VEGF receptors, protein structures affected by nintedanib (Wells, Flaherty et al. 2020), as well as the one occurring by the inhibition of Src, a molecule essential for fibroblasts activation and migration (Li, Zhao et al. 2020). To test Src inhibition in practice, we exposed TGF β -activated NHLFs to non-toxic increasing concentrations of nintedanib and A-419259, an Src inhibitor. Both molecules and more importantly A-419259 down-regulated *TKS5* and *COL1A1* expression (Figure 34B-E), reduced the number of podosomes (Figure 34F-G) and diminished aECM invasion (Figure 34H).

Last, we attempted to examine the *in vivo* therapeutic effects of A-419259. Initially, we validated that Src inhibition post to BLM treatment does decrease pulmonary fibrosis (Figure 34I). Subsequently, we treated 6 mice with the inhibitor for 6 days beginning from the 7th day post BLM. The inhibitor was administered via the respiratory route and it corresponded to 2mg/kg per animal (Figure 35A). Neither lethality nor any dramatic loss weight were witnessed (Figure 35B-C). Interestingly, pulmonary edema, inflammation and collagen production were all reduced (Figure 35D-G). In addition, collagen deposition in the lungs and the architectural disruption of the tissue, common consequence of BLM treatment, were diminished (Figure 35H-I).

To sum up, *TKS5*-mediated formation of podosomes, an active pro-fibrotic process in the PF lung. It is a druggable mechanism that can be successfully targeted by Src inhibition in an attempt to attenuate lung fibrosis via inhibition of ECM invasion.

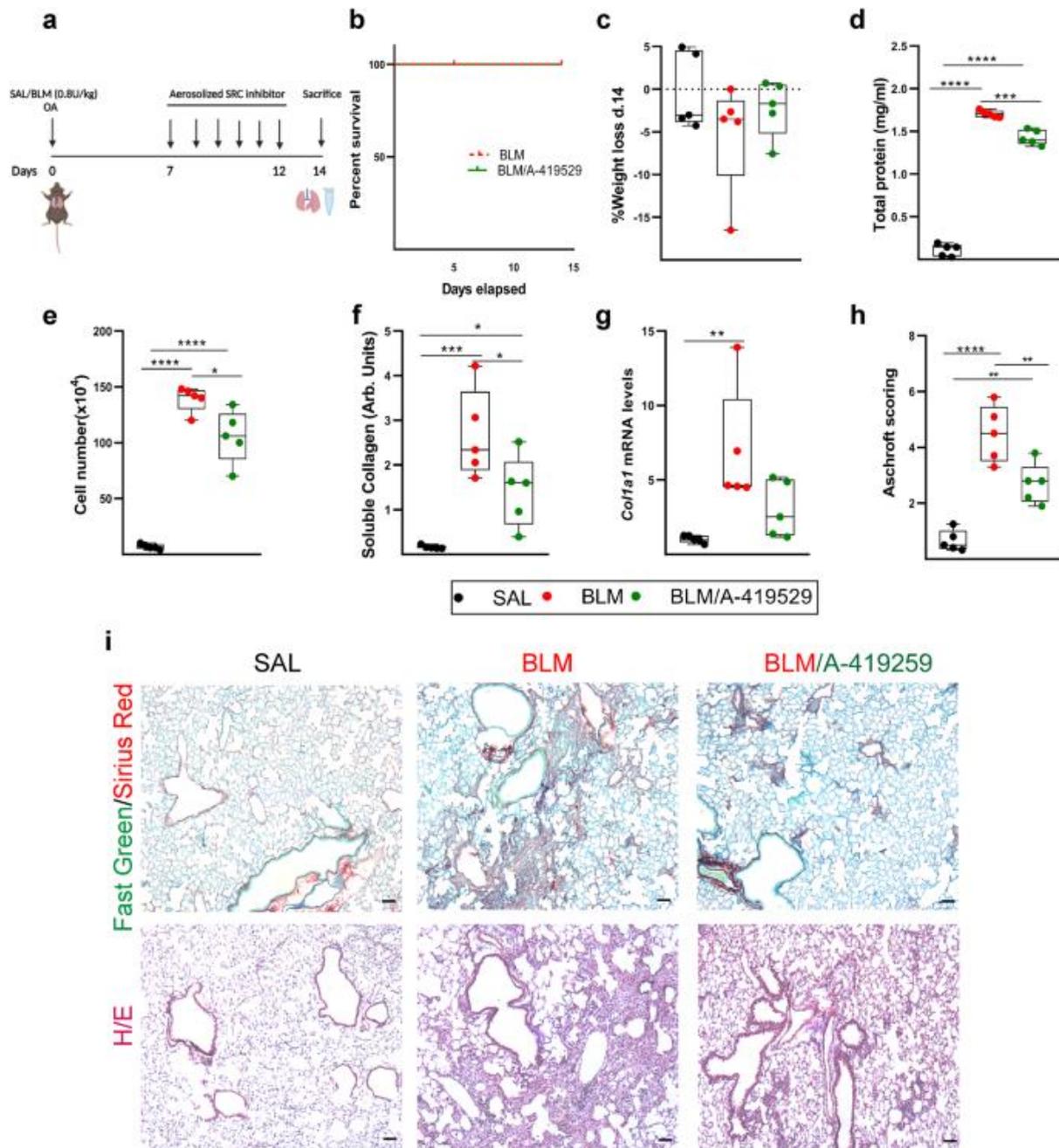


Figure 35 Inhibiting Src in the BLM-mouse model attenuates pulmonary fibrosis. **A.** BLM model and drug administration (biorender.com). **B.** Kaplan Meyer survival curve of BLM-administered mice, treated or untreated with Src-inhibitor. **C.** BLM-treated weight change of mice with or without Src-inhibition. **D.** BAL fluid protein concentration (Bradford assay; two-tailed one-way ANOVA; **** $p < 0.0001$, *** $p = 0.0001$). **E.** Number of inflammatory cells in BAL fluid (hemacytometer; two-tailed one-way ANOVA; **** $p < 0.0001$, * $p = 0.0115$). **F.** BAL fluid soluble collagen (direct red assay; two-tailed one-way ANOVA; *** $p = 0.0004$, * $p = 0.0473$). **G.** Q-RT-PCR measured *Col1a1* expression normalized over that of *B2m* and presented as fold change over control (two-tailed Kruskal Wallis; ** $p = 0.0094$). **H.** Ashcroft scoring of fibrosis load in Hematoxylin & Eosin (H/E) stained lung sections (two-tailed one-way ANOVA; **** $p < 0.0001$, ** $p = 0.0028/0.0074$). **I.** Mouse lung sections stained with Fast Green/Sirius Red and H&E; scale bars = 50 μm . Adopted from (Barbayianni, Kanellopoulou et al. 2023).

3.8 MAP3K8 against fibrosis

3.8.1 Decreased *Map3k8* expression during pulmonary fibrosis

Initially, we set to investigate *Map3k8* expression in a BLM mouse model. Towards that goal, we treated C57Bl6/J mice with BLM and sacrificed them 7 (inflammatory phase), 14 (fibrotic phase) or 21 (resolution phase) days later. Animals that have received BLM and survived until the designated sacrifice day (Figure 36A) were characterized by loss of weight and a gradual increase in pulmonary edema and tissue inflammation (Figure 36C-D).

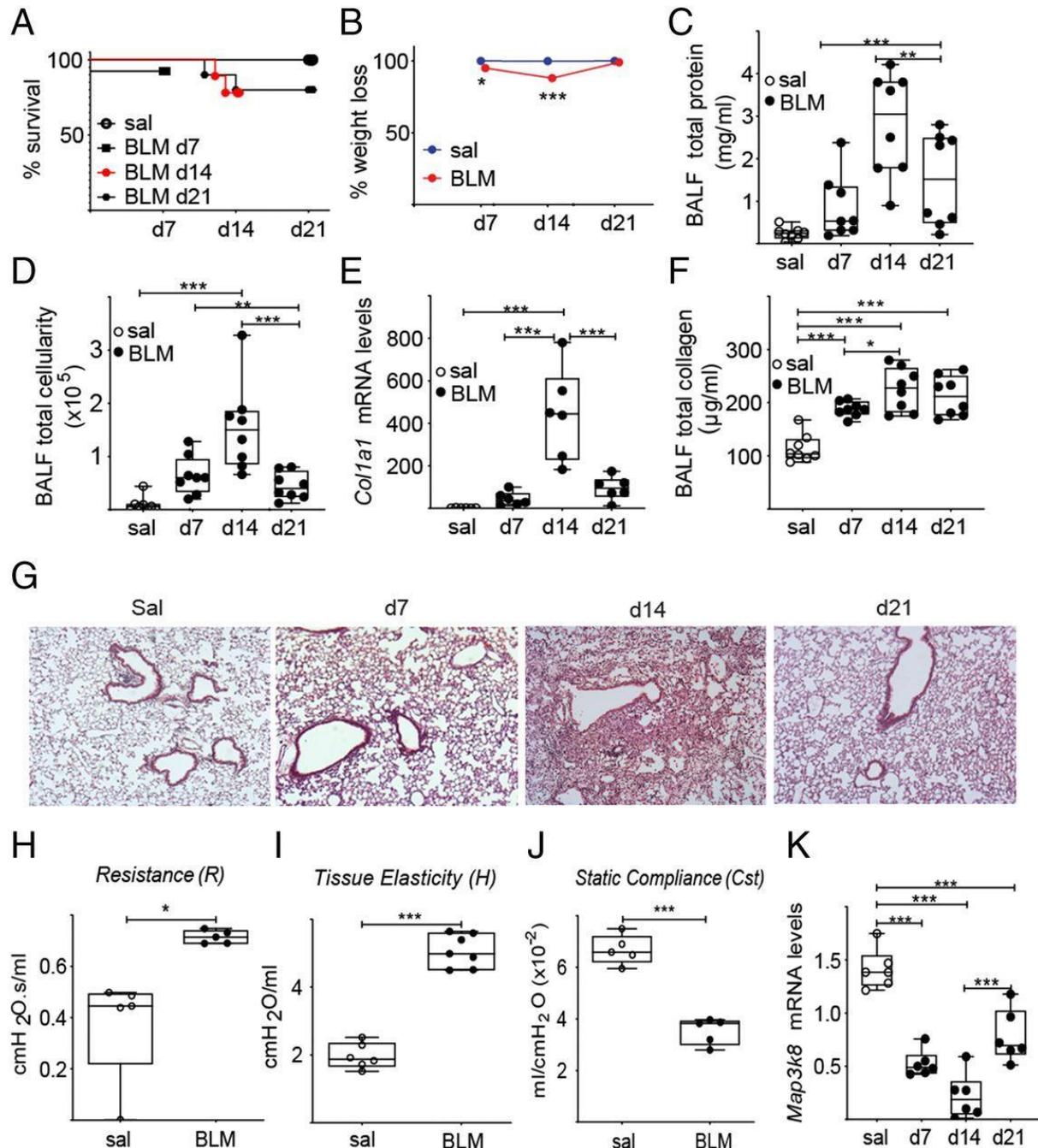


Figure 36 *Map3k8* expression is down-regulated in the BLM-induced fibrosis model. **A-B.** Survival and weight loss curves of BLM-treated/untreated C57Bl6/J mice. **C.** BAL fluid protein concentration (Bradford assay). **D.** BAL fluid inflammatory cell number (hematocytometer). **E.** Q-RT-PCR measured expression of *Col1a1* in whole-lung tissue normalized to that of *B2M*. **F.** Soluble collagen in BAL fluid (Sirius Red assay). **G.** H&E-staining of murine lungs at 7, 14, and 21 days

following BLM administration (original magnification 310). **H–J.** FlexiVent measured lung function metrics. **K.** Q-RT-PCR defined expression of *Map3k8* measured in the same samples as in E with Q-RT-PCR. Statistical significance assessed by two-way ANOVA with Bonferroni post hoc correction (A–F) and unpaired two-tailed Welch t-test (H–J); * $p = 0.05$, ** $p = 0.01$, *** $p = 0.001$. Adopted from (Zannikou, Barbayianni et al. 2021).

In parallel, we noted an over-expression of collagen (Figure 36E) and an increase of its soluble form in the BAL fluid (Figure 36F). Expansion of fibrotic regions (Figure 36G) was accompanied by an impairment of respiratory functions (Figure 36H–J). Most importantly, *Map3k8* expression was down-regulated during the inflammatory and fibrotic phase of the model compared to baseline conditions (Figure 36K).

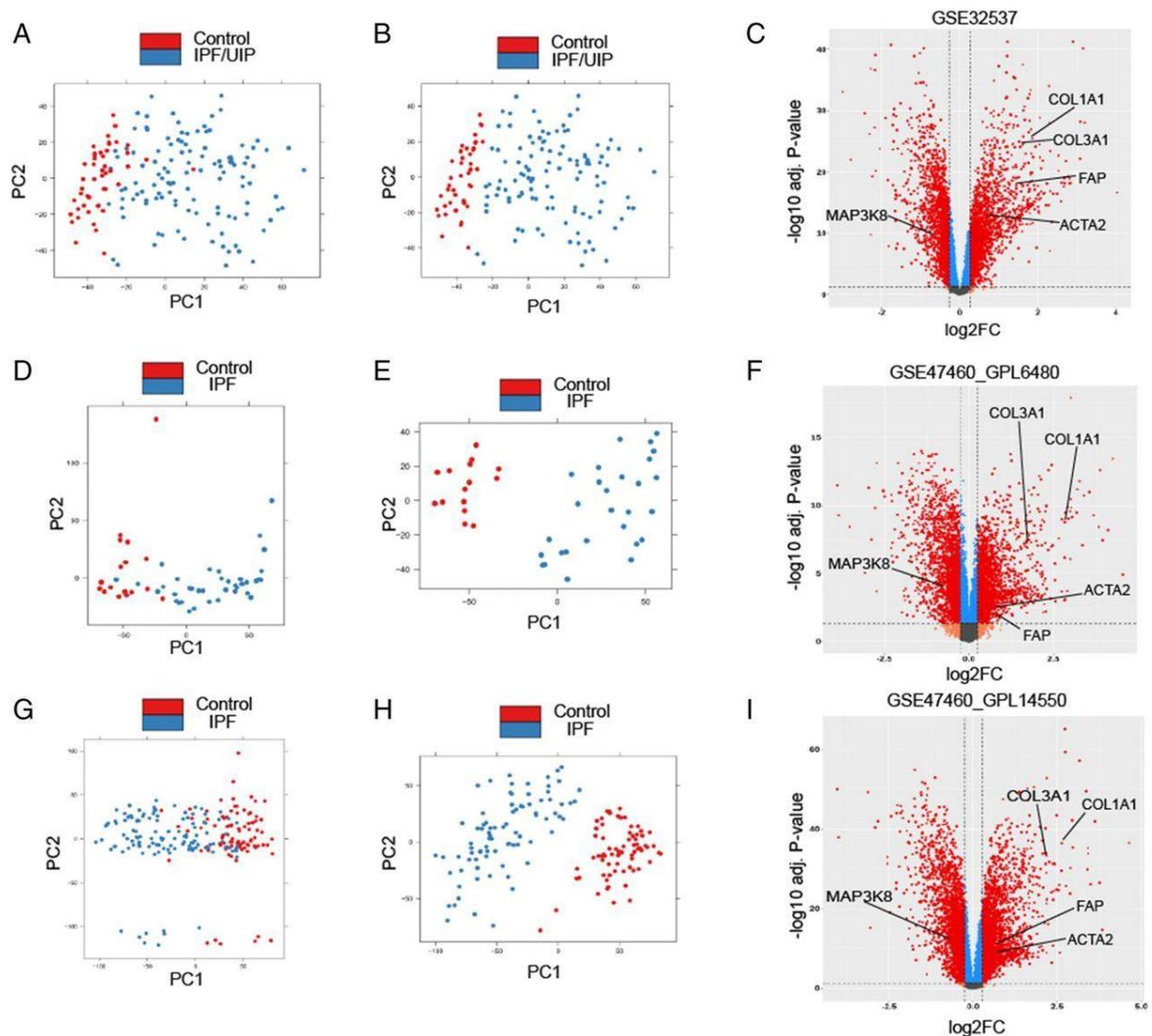


Figure 37 *MAP3K8* is under-expressed in the human fibrotic lung. **A,D,G.** Samples in reduced space prior to outlier removal. **B,E,H.** Samples in reduced space post to outlier removal. **C,F,I.** Differentially expressed genes ($|FC| \geq 1.2$ and $FDR < 0.05$) of the samples presented in (B,E,H). Adopted from (Zannikou, Barbayianni et al. 2021).

To extend our observations in IPF, we computationally re-analyzed three big cohorts of IPF and control lung samples publicly available under the GEO accession codes GSE32537 (Yang, Coldren et al. 2013) and GSE47460 (Tan, Tedrow et al. 2016). Post to background correction and rma normalization, we used PCA plots to identify and remove outlier samples (Figure 37A–

B, 37D-E, 37F-G). According to DEA results, *MAP3K8* mRNA levels were consistently reduced during pulmonary fibrosis (Figure 37C,F,I) (log2 FC and FDR corrected p-value for GSE32537, GSE47460-GPL6480, and GSE47460-GPL14550 are respectively -0.459/4.65×10⁻¹⁰; -0.658/1.06×10⁻⁰⁴; -0.668/1.46×10⁻¹³).

Motivated by previous knowledge regarding cell specific effects of *Map3k8* (Vougioukalaki, Kanellis et al. 2011), we decided to examine its expression on a cell type level. Re-analysis of a publicly available scRNA-seq dataset (GSE122960) (Reyfman, Walter et al. 2018) revealed that *MAP3K8* was expressed by both hematopoietic (HC) and non-hematopoietic cells (nHC), and primarily monocytes and macrophages (Figure S11A). DEA between macrophage subtypes underlined the under-expression of *MAP3K8* in the mostly IPF-originating SPP1+ cell cluster (FC = -1.2; FDR-corrected p = 6.11×10⁻²⁴; Figure S11B) and the aSMA+ fibroblasts (FC = -1.4; FDR-corrected p = 0.0002; Figure S11C).

In summary, *MAP3K8/Map3k8* expression is down-regulated during PF both in human and mouse lungs.

3.8.2 *Map3k8* exerts anti-inflammatory functions especially in macrophages

To dissect the role of *Map3k8* during PF, we administered BLM to a *Map3k8*^{-/-} model animals and littermate WT controls. As a result, genetically modified animals were characterized by an increased lethality rate (Figure 38A) and more severe weight loss (Figure 38B) compared to their control counterparts, while immune-related measurements (Figure 38C-D) agreed to an increased disease burden. An increase in collagen expression (Figure 38E-F) and in the number of fibrotic areas (Figure 38G) was also observed. Respiratory functions were accordingly more severely impaired in knockout than in WT animals (Figure 38H-J).

Following the lead of the *in silico* analyses revealing *MAP3K8* expression from both HC and nHC, we set to identify the relative contribution of these cellular populations to *MAP3K8* levels. For that purpose we irradiated WT and knockout mice to remove all HCs and we then proceed to bone marrow transfers with cells originating from both types of animals (Figure 39A-B). Chimeric mice were then treated with BLM and sacrificed at the fibrotic peak of the model (day 14). In agreement with upstream experimentations, *Map3k8* deficiency in HCs led to an increased lethality and weight loss (Figure 39C-D). Moreover, *Map3k8* removal in either nHCs or HCs increased lung inflammation and collagen expression (Figure 39E-H). As far as the BAL fluid is concerned, immune cells were more numerous when *Map3k8* was not-expressed by HCs (Figure 39F). Last, animals lacking *Map3k8* from their HCs were characterized by structurally deformed lungs and collagen accumulation (Figure 39I).

Motivated by the fact that *Map3k8* removal from HCs greatly affected BAL fluid immune cells in BLM-treated mice, we subsequently analyzed immune cells using an incomprehensive FACS approach (Figure 40A). At first, *Map3k8* deletion slightly affected neutrophilic infiltration (Figure 40B). Nevertheless, deletion from nHCs and especially HCs promoted macrophages accumulation. Last, we observed accumulation of lymphocytes when HC, but not nHCs, were *Map3k8*-deficient (Figure 40C).

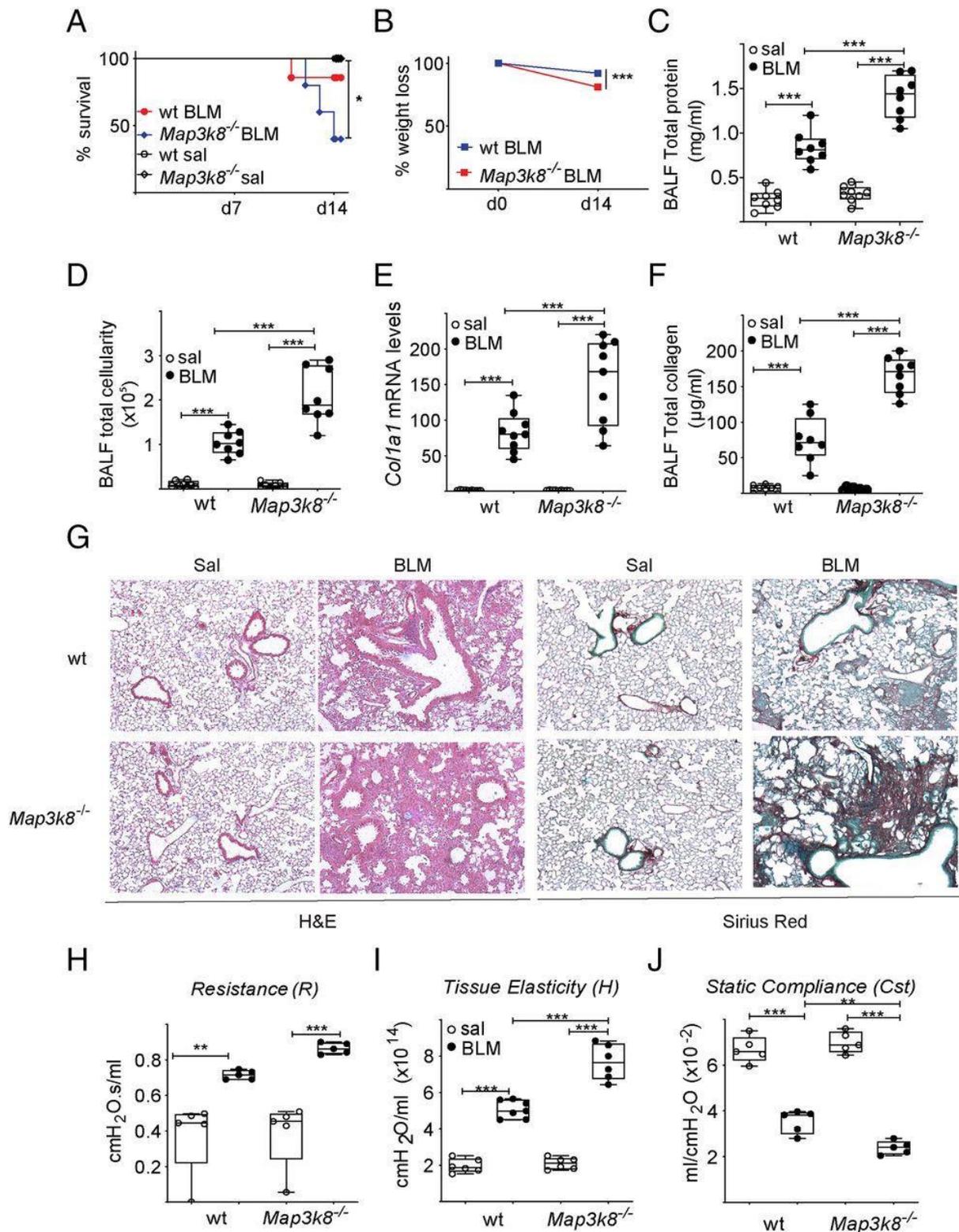


Figure 38 *Map3k8* deficiency induces fibrosis establishment. **A-B.** Survival and weight loss curves of mice treated with BLM or saline. **C.** BAL fluid protein concentration (Bradford assay). **D.** Number of inflammatory cells BAL fluid (hematocytometer). **E.** Q-RT-PCR measured expression of *Col1a1* in whole-lung tissue normalized to that of *B2M*. **F.** Soluble collagen levels of the BAL fluid (Sirius Red assay). **G.** H&E- and Sirius Red- staining of murine lungs (original magnification 310). **H-J.** FlexiVent assessed respiratory function metrics. Statistical significance was assessed with two-way ANOVA

followed by Bonferroni post hoc correction; *p = 0.05, **p = 0.01, ***p = 0.001. Adopted from (Zannikou, Barbayianni et al. 2021).

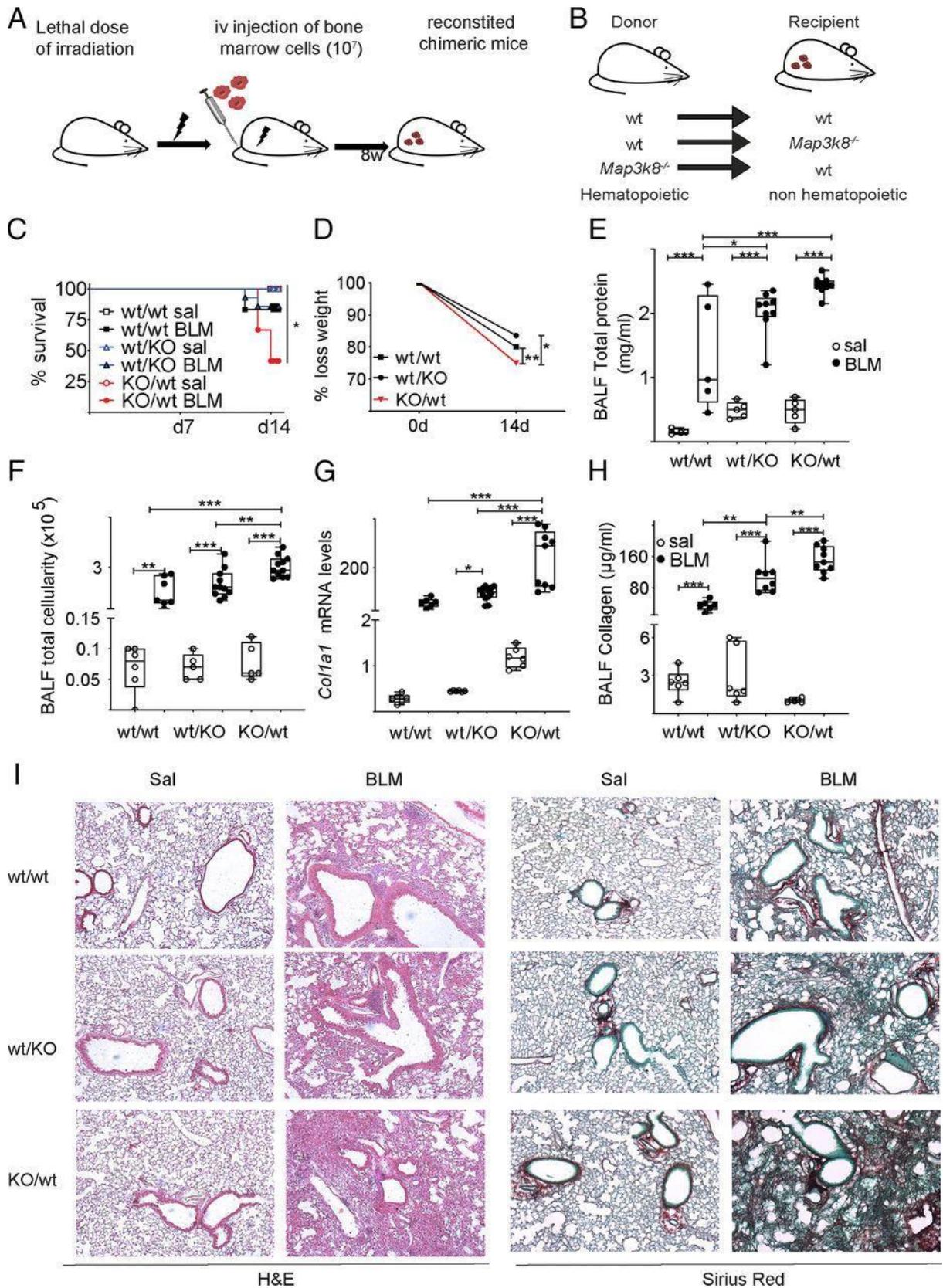


Figure 39 *Map3k8* deletion from hematopoietic or non-hematopoietic lineage cells intensifies lung inflammation and fibrosis. Chimeric mice having HCs with the donor’s genetic background (WT or

Map2k8^{-/-}) were treated with BLM. Fibrosis was evaluated 14 days post BLM administration. **A.** Depiction of bone marrow transfers. **B.** Chimeric mice obtained from the bone marrow transfers presented in (A). **C-D.** Survival and weight loss curves post BLM treatment. **E.** BAL fluid protein concentration (Bradford assay). **F.** BAL fluid Inflammatory cell numbers (hematocytometer). **G.** Q-RT-PCR measured expression of *Colla1* in whole-lung tissue normalized to that of *B2M*. **H.** BAL fluid soluble collagen (Sirius Red assay). **I.** H&E- and Sirius Red murine lungs stainings (original magnification 310). Statistical significance evaluated by two-way ANOVA followed by Bonferroni post hoc correction (* $p = 0.05$, ** $p = 0.01$, *** $p = 0.001$). Adopted from (Zannikou, Barbayianni et al. 2021).

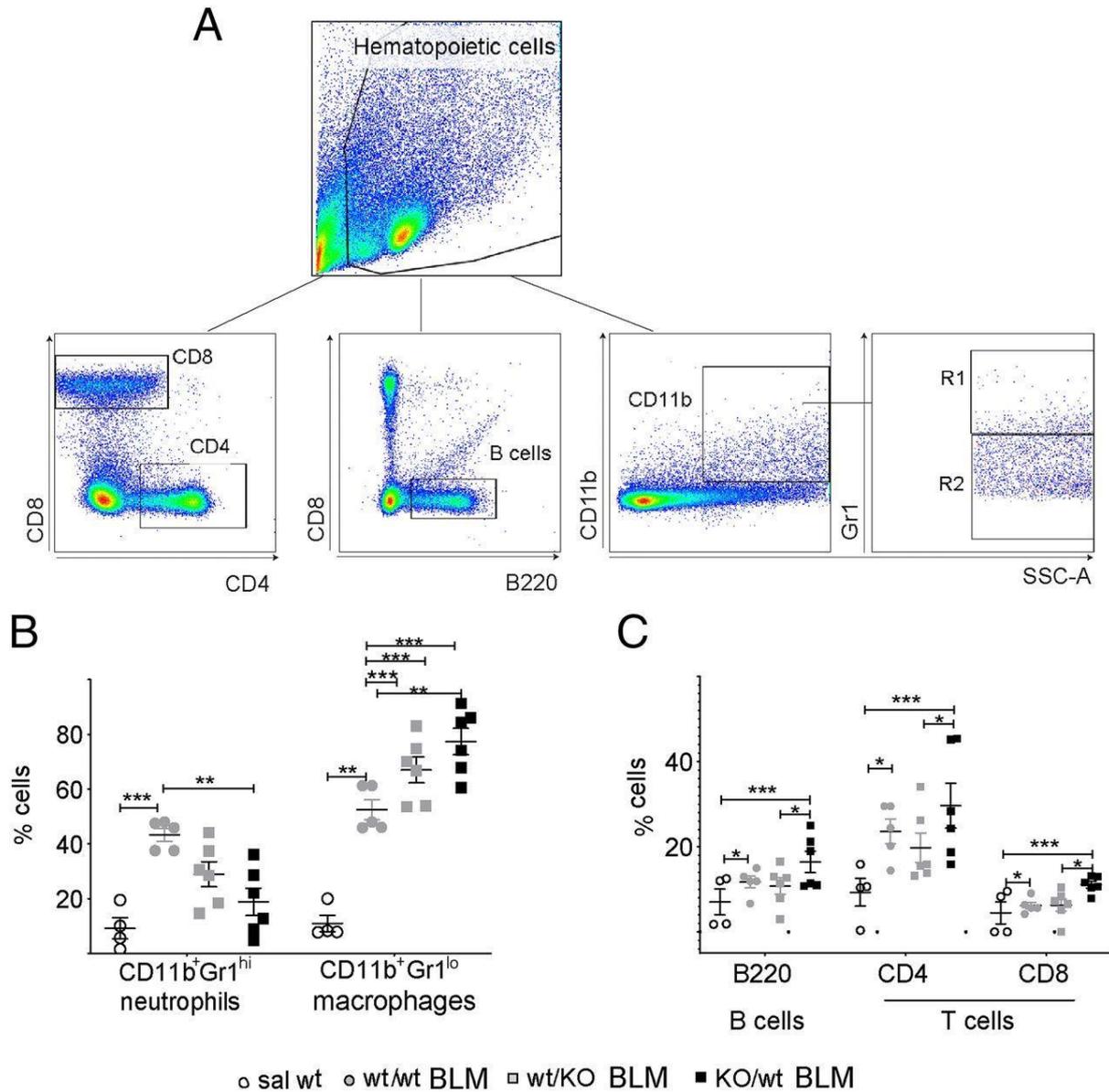


Figure 40 MAP3K8 deficiency increases the number of inflammatory cells in the BAL fluid of BLM-treated animals. **A.** FACS gating strategy. **B-C.** FACS analysis of the indicated cell types (two-way ANOVA with Bonferroni post hoc correction; * $p = 0.05$, ** $p = 0.01$, *** $p = 0.001$). Adopted from (Zannikou, Barbayianni et al. 2021).

Due to the above described reduced *Map3k8* expression in IPF macrophages and the exacerbation of inflammation/fibrosis as a result of its deficiency, we decided to examine the role of *Map3k8* in macrophage homeostasis. To do so, we created a mouse strain

(*LysMMap3k8^{-/-}*) whose macrophages and granulocytes lacked *Map3k8*. These transgenic mice were treated with BLM and compared to BLM-treated WT littermates. It was observed that *LysMMap3k8^{-/-}* animals suffered from severe fibrosis reflected in all performed measurements (Figure 41). More importantly, *Map3k8*-deficient BAL fluid macrophages had the same phenotype with the *Map3k8*-deficient HCs (Figure 42; Figure 39) suggesting that both *Map3k8* and lung macrophages play an important role in PF progression.

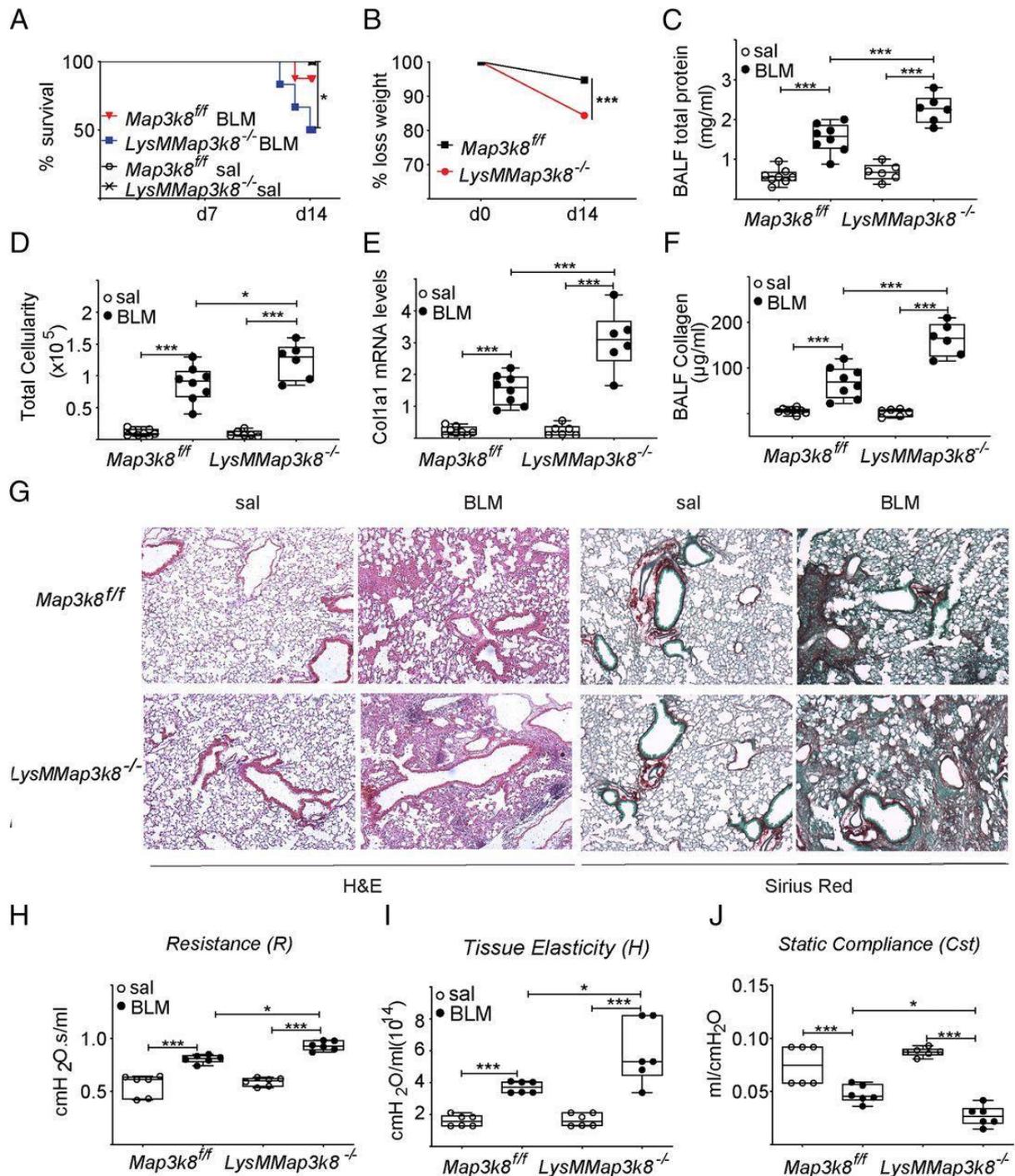


Figure 41 *Map3k8* deletion from LysM+ cells intensifies pulmonary inflammation and fibrosis. *LysMMap3k8^{-/-}* mice lacking *Map3k8* in macrophages/granulocytes and wt littermates were treated with BLM. Fibrosis establishment was evaluated 14d post treatment. A-B. Survival and weight loss curves of mice treated with BLM or saline. C. BAL fluid protein concentration (Bradford assay). D. BAL fluid

inflammatory cell numbers (hematocytometer). **E.** Q-RT-PCR measured expression of *Coll1a1* in whole-lung tissue normalized to that of *B2M*. **F.** BAL fluid soluble collagen (Sirius Red assay). **G.** H&E- and Sirius Red murine lungs stainings (original magnification 310). **H–J.** FlexiVent defined metrics of proper respiratory functioning. Statistical significance was tested with two-way ANOVA followed by Bonferroni post hoc correction (* $p = 0.05$, *** $p = 0.001$). Adopted from (Zannikou, Barbayianni et al. 2021).

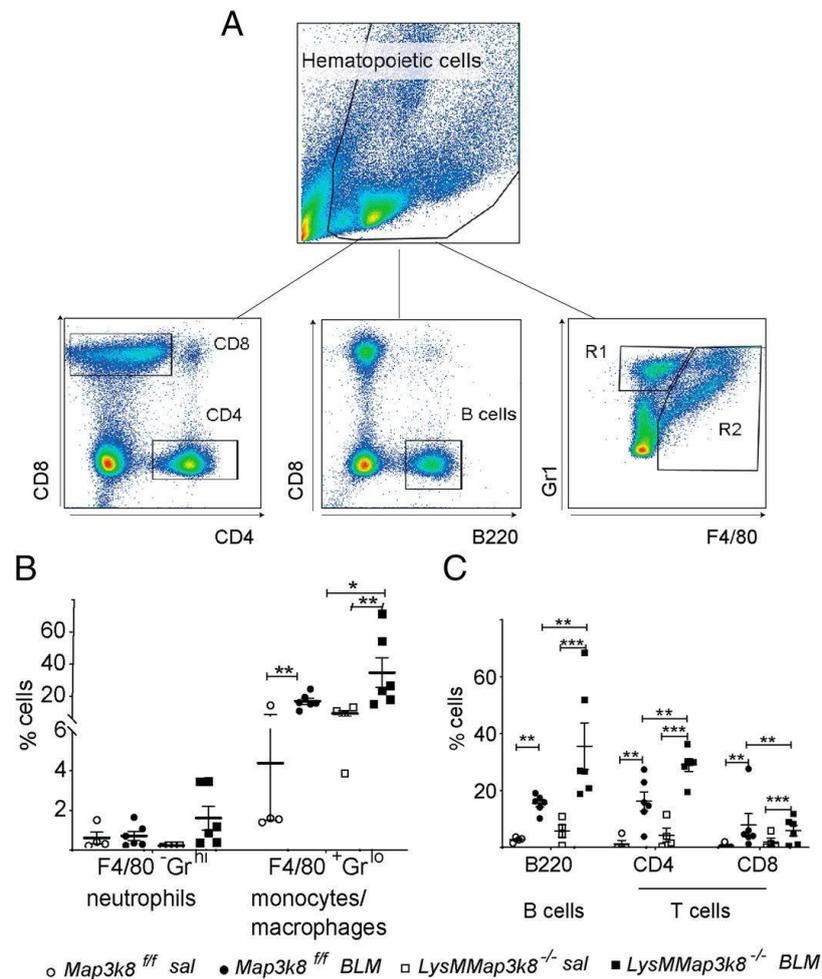


Figure 42 *Map3k8* deficiency from *LysM*⁺ cells increases the number of inflammatory cells in the BAL fluid of BLM-treated mice. **A.** FACS analysis gating strategy. **B–C.** FACS analysis results. Statistical significance was assessed with two-way ANOVA followed by Bonferroni post hoc correction (* $p = 0.05$, ** $p = 0.01$, *** $p = 0.001$). Adopted from (Zannikou, Barbayianni et al. 2021).

3.8.3 *Map3k8*-produced PGE₂ attenuates pulmonary fibrosis

Prostaglandin E₂ (PGE₂) production from arachidonic acid (AA) via regulation of Cox-2 levels is known to be regulated by *Map3k8* (Figure 43A) (Roulis, Nikolaou et al. 2014, Roulis, Kaklamanos et al. 2020). PGE₂ is an eicosanoid that has been proposed to act in an anti-fibrotic function in the lungs (Bozyk and Moore 2011).

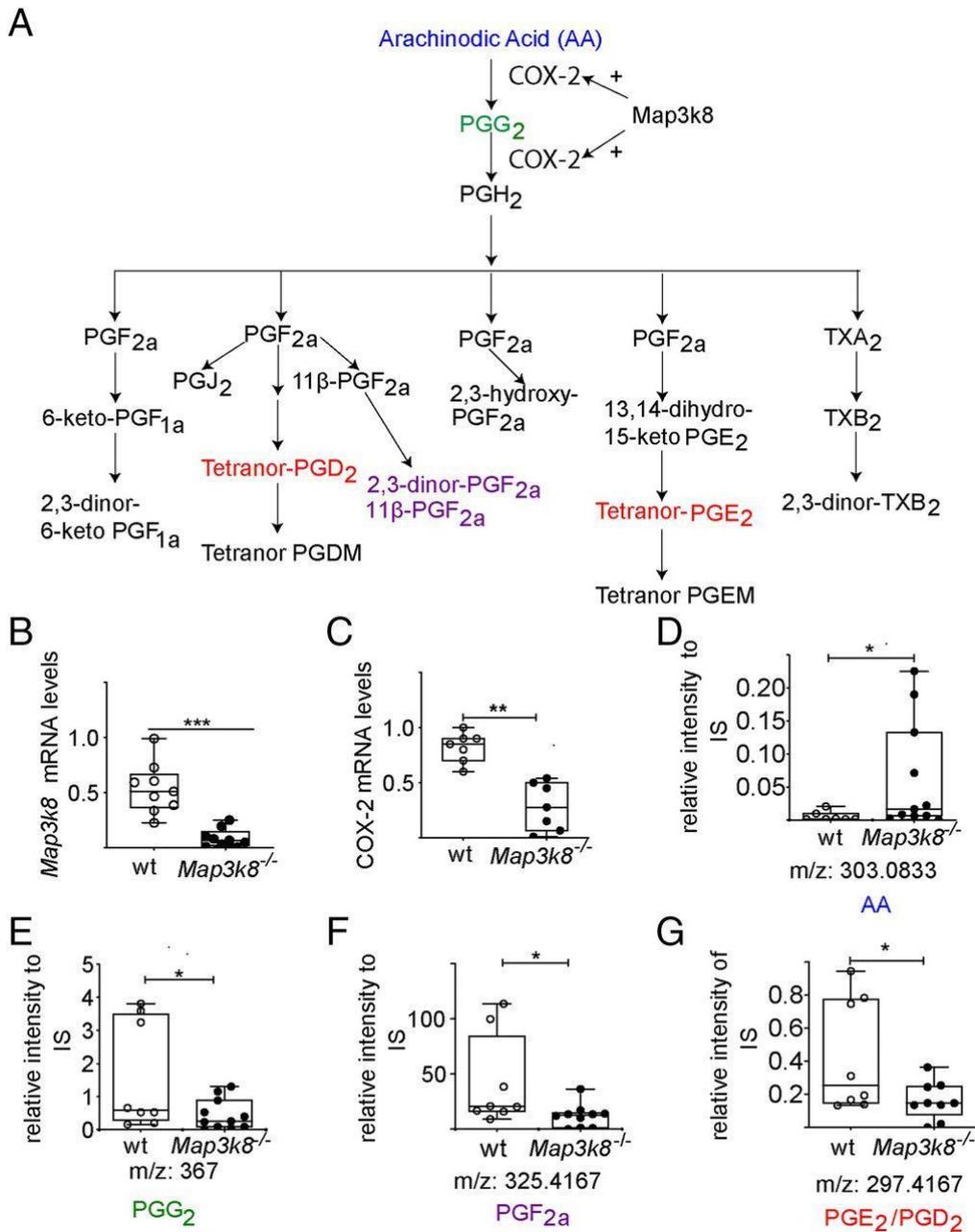


Figure 43 Map3k8 modifies the Cox-2-AA-PGE2 axis of the lungs. **A.** Schematic representation of AA metabolism; measured lipid species are color coded. **B-C.** Map3k8 and Cox-2 mRNA levels in whole-lung tissue of the indicated mice and treatments were determined with Q-RT-PCR analysis; values were normalized to the expression of B2M. **D-G.** Tandem mass spectrometry determination of AA and PGE2 metabolites in BALFs. Statistical significance was evaluated using unpaired two-tailed Welch t-test (* $p = 0.05$, ** $p = 0.01$, *** $p = 0.001$). Adopted from (Zannikou, Barbayianni et al. 2021).

Thus, we were interested in assessing the effects of *Map3k8* deletion (Figure 43B) on PGE2 production during fibrosis. Examining BLM-treated mice at the peak of their fibrotic phenotype we observed that *Cox-2* expression was decreased (Figure 43C), similarly to that of *Map3k8* removal from HCs and non-HCs (Figure S12A-B), as well as to that of macrophages (Figure S12C-D). These results validate *Map3k8* expression from both nHCs and HCs along with the regulation of *Cox2* transcription from *Map3k8*. In addition, measurement of AA (Figure 43D) and PGE2 levels (Figure 43E-G) in *Map3k8*^{-/-} mice, revealed an impaired AA metabolism.

Similar observations were made in chimeric mice (Figure S13), suggesting *Map3k8* as a crucial regulator of PGE2 abundance in the lungs.

Last, in an attempt to establish a connection between PGE2 and PF, we administered a stable PGE2 analogue (dmPGE2) to WT and *Map3k8*^{-/-} mice post to BLM-treatment. PGE2 restoration restrained the intense fibrotic events observed in *Map3k8*^{-/-} animals, while at the same time reduced fibrosis severity in their WT littermates (Figure 44). Interestingly, treatment with PGE2 limited the inflammatory influx that characterizes fibrosis progression in the IPF BLM-based animal model (Figure 45).

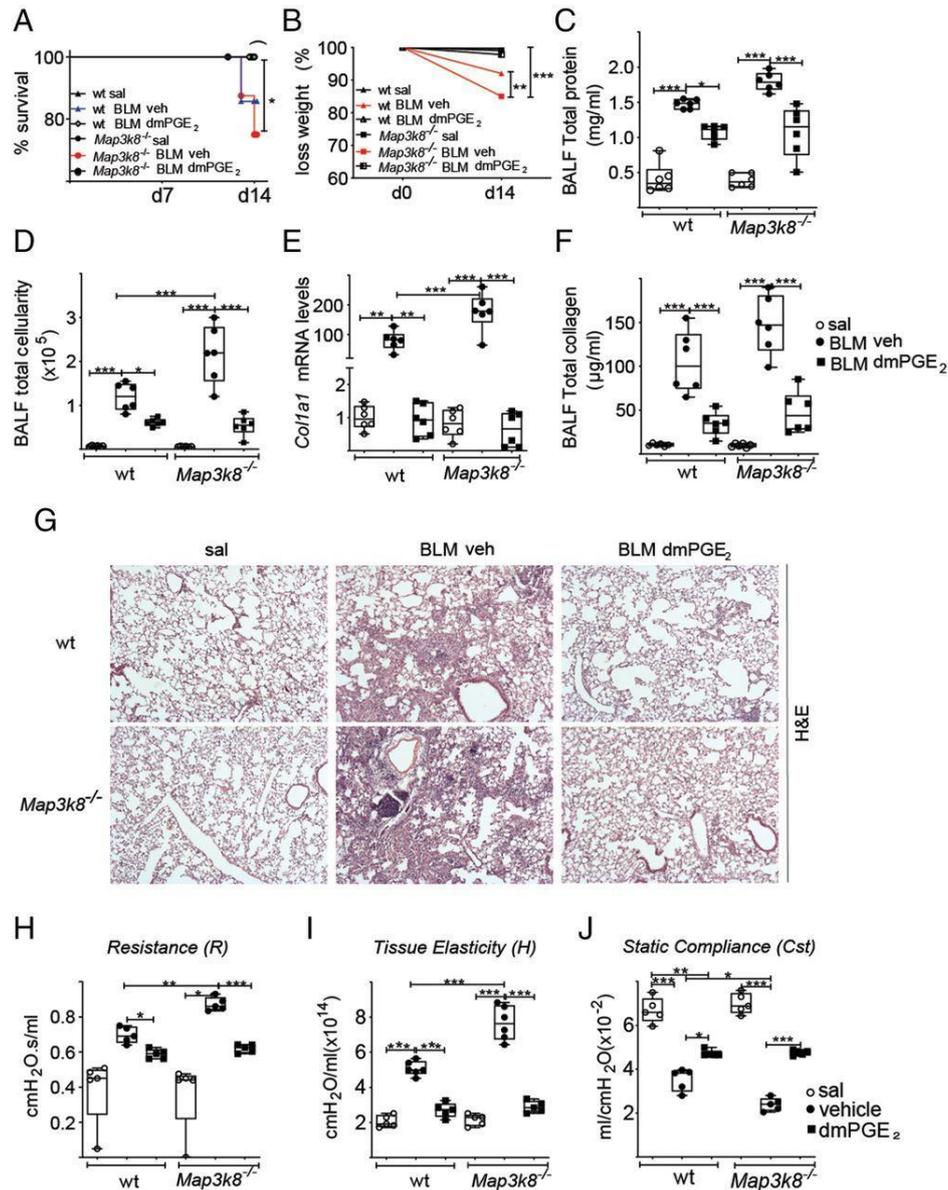


Figure 44 PGE2 reverses the *Map3k8*-established profibrotic impact and inhibits the progression of pulmonary fibrosis and inflammation. **A-B**. Survival and weight loss curves of mice treated with BLM or saline. **C**. BAL fluid protein concentration (Bradford assay). **D**. BAL fluid Inflammatory cell numbers (hematocytometer). **E**. Q-RT-PCR measure expression of *Col1a1* in whole-lung tissue normalized to that of *B2M*. **F**. BAL fluid soluble collagen (Sirius Red assay). **G**. Murine lung sections stained with H&E (original magnification 310). **H-J**. FlexiVent assessed metrics of pulmonary function

Statistical significance was evaluated with two-way ANOVA followed by Bonferroni post hoc correction (*p = 0.05, **p = 0.01, ***p = 0.001). Adopted from (Zannikou, Barbayianni et al. 2021).

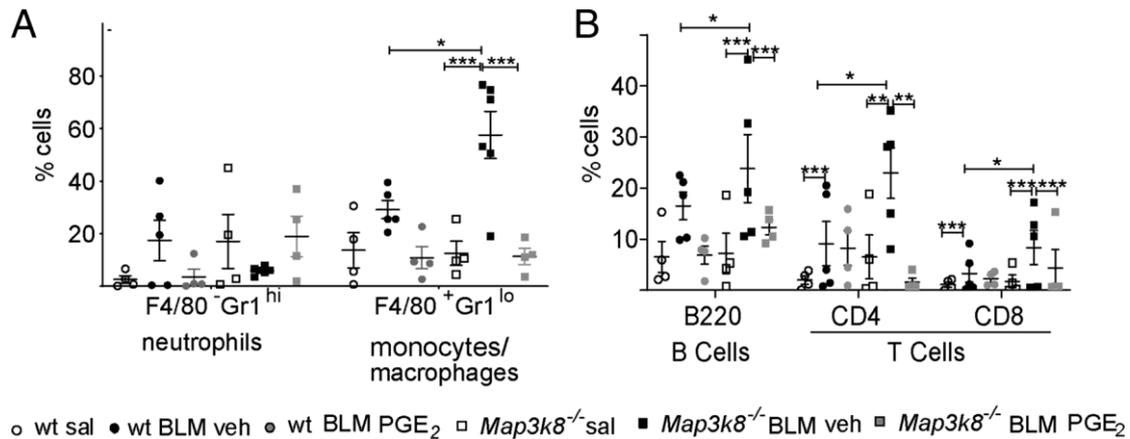


Figure 45 PGE2 decelerates the accumulation of inflammatory cells in mouse lungs. A-B. FACS analysis BAL fluid cells. The same gating strategy as above was used. Statistical significance was assessed with two-way ANOVA followed by Bonferroni multiple comparisons test (*p = 0.05, **p = 0.01, ***p = 0.001). Adopted from (Zannikou, Barbayianni et al. 2021).

In conclusion, *Map3k8* function is crucial for lung homeostasis maintenance. During fibrosis, *Map3k8* anti-fibrotic effects are impaired via down-regulation of its expression, which in turn inhibits Cox-2 mediated production of PGE2.

3.9 Dendritic cells and COVID-19

During the worldwide COVID-19 pandemic, there were several indications regarding the underlying commonalities of PF and SARS-CoV-2 infection (Bharat, Querrey et al. 2020, Drake, Docherty et al. 2020, Osuchowski, Winkler et al. 2021). Motivated by these preliminary reports, we set to examine *ENPP2* potential implication in COVID-19.

3.9.1 Increase in *ENPP2* levels during COVID-19

Working towards that goal, we initially examined *ENPP2* mRNA levels via wet laboratory methods. Briefly, measurement of *ENPP2* mRNA levels from nasopharyngeal swabs via Q-RT-PCR (Table 15) revealed that mild and severe COVID-19 patients had increased transcript numbers compared to control individuals (Figure 46). Subsequently, we used ELISA to quantify ATX levels in the serum of an independent set of COVID-19 patients stationed at the Evangelismos hospital in Athens, Greece. The cohort was comprised of individuals admitted to either WARD or intensive care unit (ICU), with the latter receiving or not dexamethasone (Dex) treatment (Table 16). ICU patients not receiving Dex had increased levels of ATX in their circulation compared to WARD patients, implicating *ENPP2* in disease severity (Figure 47A). Nevertheless, no significant correlation was observed between ATX concentration and severity markers listed in Table 16. Similarly, not any significant ATX change was related to patients' sex or comorbidities (Figure S14).

Subsequently, we set to examine the factors potentially regulating this catholic increase in ATX abundance. Initially, we focused on IL-6, the molecule most predictive of a COVID-19-related cytokine storm (Ulhaq and Soraya 2020, Zhang, Hou et al. 2020), as it has also been reported to stimulate ATX expression (Castelino, Bain et al. 2016, Sun, Wang et al. 2017).

Table 15 *ENPP2* transcription is up-regulated in the nasopharynx of COVID-19 patients compared to healthy counterparts. Statistical significance was evaluated with one-way ANOVA

followed by Bonferroni post-hoc correction (**** p < 0.0001). Adopted from (Nikitopoulou, Fanidis et al. 2021).

	Negative Healthy	Positive Mild	Positive Severe/Critical
Number of patients (n)	21	21	21
ATX ($2^{-\Delta\Delta C_t}$, mean \pm SD)	2.15 \pm 1.37	5.38 \pm 2.34 ****	5.76 \pm 2.18 ****
Sex			
Male	8 (38.09%)	10 (47.6%)	17 (80.95%)
Female	13 (61.9%)	9 (42.8%)	4 (19.04%)
Not recorded	0	2 (9.52%)	0
Age (years, mean \pm SD)	50.15 \pm 20.86	37.78 \pm 11.89	63.38 \pm 17.23

Table 16 Clinical, epidemiological and laboratory characteristics of Evangelismos general hospital COVID-19 patients. Statistical significance was evaluated with one-way ANOVA followed by Bonferroni post-hoc correction (**** p < 0.0001). Adopted from (Nikitopoulou, Fanidis et al. 2021).

	WARD No Dex	ICU No Dex	ICU + Dex
Number of patients (n)	47	37	32
ATX (ng/mL, mean \pm SD)	310.32 \pm 98.85 *	443 \pm 172.90	246.15 \pm 73.74 *
Sex			
Male	33 (70.21%)	31 (83.78%)	22 (62.5%)
Female	14 (29.78%)	6 (16.21%)	10 (31.25%)
Age (years, mean \pm SD)	54.63 \pm 15.46	63.54 \pm 10.89	65.5 \pm 10.7
Comorbidities n (%)			
Hypertension	13 (27.65%)	17 (45.94%)	12 (37.5%)
Diabetes	4 (8.51%)	5 (13.51%)	5 (15.62%)
Coronary artery disease	8 (17.02%)	4 (10.81%)	4 (12.5%)
COPD	1 (2.12%)	1 (2.7%)	2 (6.25%)
Asthma	2 (4.25%)	1 (2.7%)	1 (3.12%)
Hyperlipidemia	9 (19.14%)	9 (24.32%)	8 (25%)
Hepatitis	0 (0%)	1 (2.7%)	0 (0%)
COVID-19 treatment	0	11	
Azithromycin/chloroquine/lopinavir/ritonavir	6	7	
Azithromycin/chloroquine	0	2	
Lopinavir/ritonavir/chloroquine	0	3	
Chloroquine Plasma	0	1	
Clinical measurements			
Mean arterial pressure (mmHg)	83.19 \pm 8.86	82.83 \pm 16.52	77.55 \pm 8.54
PaO ₂ /FiO ₂ (mmHg)	301.5 \pm 79.81 *	194.86 \pm 86.64	85.94 \pm 15.97 *
Glucose (mg/dL)	133.5 \pm 113.3	164.53 \pm 77.73	164.06 \pm 75.40
Creatinine (mg/dL)	0.9 \pm 0.33	1.02 \pm 0.32	0.95 \pm 0.72
CRP (mg/dL)	6.8 \pm 8.96	14 \pm 10.17	13.83 \pm 9.6
Total bilirubin (mg/dL)	0.5 \pm 0.33	0.73 \pm 0.5	0.61 \pm 0.29
White blood cell count (per μ L)	6995 \pm 3468	10,125 \pm 4633	11,705 \pm 10,372
Neutrophils (%)	69.34 \pm 13.51	81.34 \pm 6.64	83.12 \pm 12.2
Lymphocytes (%)	24.03 \pm 10.89 *	12.63 \pm 5.63	11.12 \pm 11.23
Platelets (per μ L)	240,297 \pm 110,028	237,783 \pm 101,338	257,000 \pm 79,581
INR (median IQR)	1.06 \pm 0.09	2.07 \pm 5.73	1.26 \pm 0.65
D-dimer (pg/mL)	1.19 \pm 1.72	0.47 \pm 0.26	1.39 \pm 0.93
AST (IU/L)	36.65 \pm 30.65	54.18 \pm 39.95	121.4 \pm 329.9
ALT (IU/L)	33.15 \pm 23.58	45.9 \pm 28.08	60.8 \pm 72.4
LDH (U/L)	286.36 \pm 122.08 *	498.48 \pm 242.34	591.23 \pm 490.84
Fibrinogen (mg/dL)	514.06 \pm 176.18	638.18 \pm 158.76	630.3 \pm 172.2
Ferritin (pg/mL)	513.48 \pm 815.55	2786 \pm 694.48	912.47 \pm 826.91
APACHE II score	5.25 \pm 2.94	14.27 \pm 5.08	15.4 \pm 3.89
SOFA score	2 \pm 1	6.83 \pm 3.08	5.4 \pm 1.81

Spearman's correlation revealed a significant relationship between IL-6 and ATX levels (Figure 47B). In addition, following the higher abundance molecules reported by a previous publication (Vassiliou, Keskinidou et al. 2021) on a subset of the Evangelismos cohort, we discovered that ATX levels positively correlated with those of endothelium proteins (Figure S15), thus implicating the former in COVID-19-related endothelium impairment.

Last, as Dex is used for the treatment of lung inflammatory conditions (Xu, Qiao et al. 2009, Patil, Naveen Kumar et al. 2018), such as ARDS, and has also proven beneficial for certain

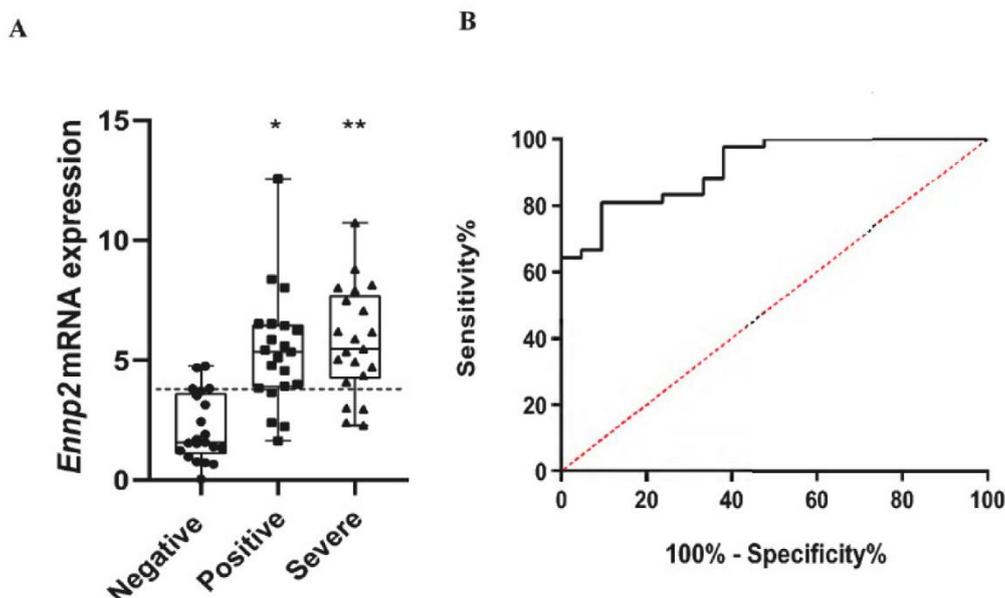


Figure 46 ENPP2 transcripts are increased in nasopharynx swabs of COVID-19 patients. **A.** ENPP2 expression in two patient and a control individual group (dotted line marks the optimal threshold value). **B.** ROC curve generated post to patient groups merging. (* $p = 4 \times 10^{-6}$; ** $p = 2.92 \times 10^{-7}$). Adopted from (Nikitopoulou, Fanidis et al. 2021).

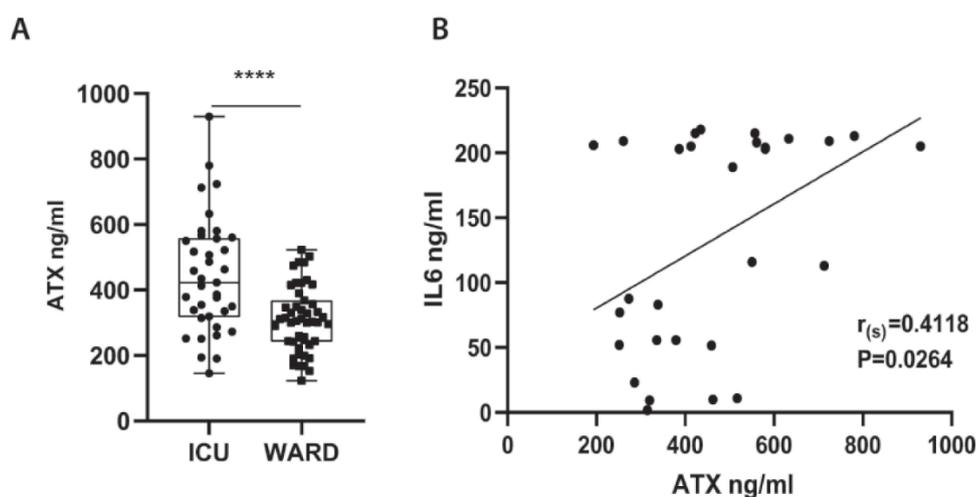


Figure 47 ATX levels are increased in the serum of ICU admitted patients and correlate to IL-6 levels. **A.** ELISA-measured levels of ATX from the serum of Evangelismos patients that had not been treated with Dex (unpaired t-test; **** $p < 0.0001$). **B.** ATX levels in the serum correlated to those of IL-6 (Spearman's correlation). Adopted from (Nikitopoulou, Fanidis et al. 2021).

COVID-19 patients (Sterne, Murthy et al. 2020, The Recovery Collaborative Group 2020), we wanted to investigate its impact on ATX serum abundance. Interestingly, intubated Dex-treated patients had decreased amounts of circulating ATX compared to the intubated untreated ones, while no difference was witnessed between patients not receiving any mechanical ventilation (Figure 48A). The same observations (Figure 48B) were made from the examination of another independent cohort hospitalized in the Patras University hospital (Table 17). In both cases, artificially ventilated Dex-untreated patients had the highest amounts of serum ATX. More importantly, ATX levels varied significantly between differential survival patient groups, with deceased untreated patients having the highest ATX levels (Figure 48C-D).

Table 17 Clinical, epidemiological and laboratory characteristics of University Hospital of Patras ICU patients. Statistical significance was evaluated with one-way ANOVA followed by Bonferroni post-hoc correction (** p < 0.05). Adopted from (Nikitopoulou, Fanidis et al. 2021).

	ICU No Dex	ICU + Dex
Number of patients (n)	12	23
ATX (ng/mL, mean \pm SD)	624.36 \pm 203.5	404.16 \pm 145.5 **
Sex		
Male	9 (75%)	18 (78.26%)
Female	3 (25%)	5 (21.73%)
Age (years, mean \pm SD)	66.75 \pm 13.31	59.43 \pm 15.42
Comorbidities n (%)		
Hypertension	5 (41.6%)	10 (43.47%)
Diabetes	0 (0%)	4 (17.39%)
Coronary artery disease	2 (16.6%)	1 (4.34%)
COPD	0 (0%)	2 (8.69%)
Asthma	0 (0%)	0 (0%)
Hyperlipidemia	3 (25%)	6 (26.08%)
Hepatitis	(%)	(%)
COVID 19 treatment		
Azithromycin/chloroquine/lopinavir/ritonavir	4 (33.33%)	0 (0%)
Azithromycin/chloroquine	6 (50%)	5 (21.73%)
Lopinavir/ritonavir/chloroquine	1 (8.33%)	0 (0%)
Chloroquine	0 (%)	0 (%)
Plasma	0 (%)	0 (%)
Clinical measurements		
Glucose (mg/dL)	120.5 \pm 24.57	154.04 \pm 47.98 *
Creatinine (mg/dL)	0.95 \pm 0.42	0.93 \pm 0.38
CRP (mg/dL)	8.53 \pm 5.51	15.55 \pm 12.68
Total bilirubin (mg/dL)	0.67 \pm 0.27	0.7 \pm 0.36
Lymphocytes (absolute number)	0.62 \pm 0.35	0.83 \pm 0.5
INR (median IQR)	1.09 \pm 0.13	1.07 \pm 0.12
D-dimer (pg/mL)	2.21 \pm 2.14	1.45 \pm 1.72
LDH (U/L)	370 \pm 129	455.08 \pm 188.56
Fibrinogen (mg/dL)	630.55 \pm 168.1	545 \pm 182.32
Ferritin (pg/mL)	950 \pm 382.64	1131.09 \pm 1223.55

In conclusion, *ENPP2* transcription and translation do up-regulate in COVID-19 patients in parallel to several important molecules, including IL-6. Moreover, the therapeutic mechanism of Dex seems to be reflected and/or manifested via ATX reduction in the blood.

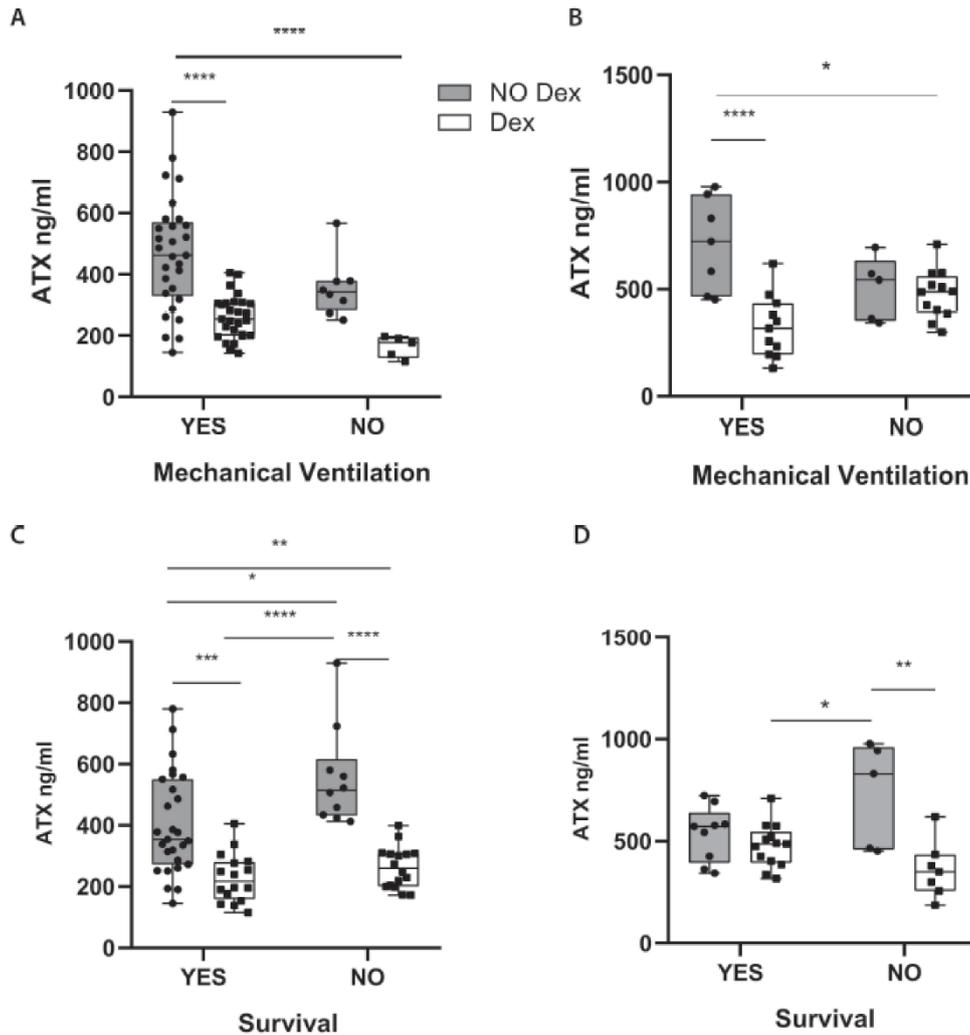


Figure 48 ATX quantity in the serum of hospitalized patients is affected by dexamethasone treatment. ELISA measured ATX levels in the serum of COVID-19 patients admitted in the ICU of the Evangelismos (A,C) or the Patras University hospital (B,D). The measurements in the Dex groups in (A) and (C), are the same as with those in Figure 47A. Statistical significance was evaluated with 2-way ANOVA followed by Bonferroni post hoc correction (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and **** $p < 0.0001$). Adopted from (Nikitopoulou, Fanidis et al. 2021).

3.9.2 Cellular pattern of ENPP2 expression in COVID-19

In an attempt to identify the cellular source of *ENPP2* during SARS-CoV-2 infection, we have re-analyzed RNA-seq data of single cells isolated from the circulation, the upper and the lower respiratory track of COVID-19 patients (Table S19).

Beginning from the two nasopharyngeal datasets, *ENPP2* was detected in immune cells and specifically in macrophages and NK cells (Figure 49A; Figure S16A). In the PBMC-focusing circulation datasets, *ENPP2* was primarily expressed by plasmacytoid dendritic cells (pDCs) (Figure 49B; Figure S16B). The same cell type was also witnessed as the main source of *ENPP2* transcripts among human BAL fluid cells (Figure 49C; Figure S16C). In the lung, *ENPP2* was produced by mesothelial, arterial and monocytic group cells (Figure 49D; Figure S16D). A similar pattern of expression was revealed in the lung of IPF patients (Figure S16E; Table S19), indicating a potentially common mechanism of action between the two diseases.

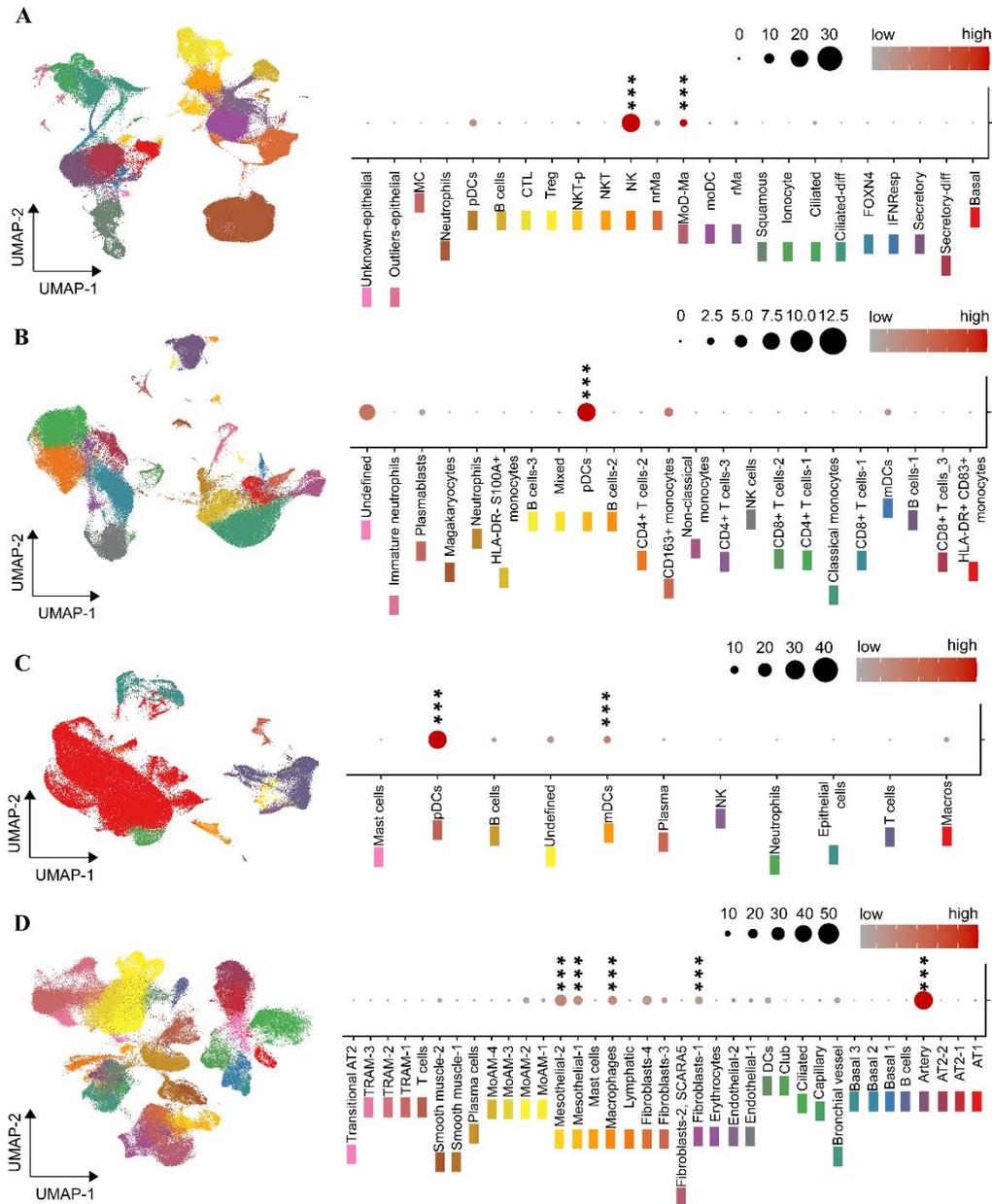


Figure 49 *ENPP2* expression pattern across tissues and cell types during COVID-19. **A.** Nasopharyngeal swabs (PMID: 32591762). **B.** PBMC (PMID: 32810438). **C.** BAL fluid (PMID: 32398875). **D.** lung tissue (PMID: 33257409). Dimensionality reduction plots iconize each site's cellular composition. Dot plots depict *ENPP2* expression pattern relative to the cell types of the respective UMAP plot (color coded). Cell types marked by *ENPP2* expression are highlighted with the use of stars (Wilcoxon rank sum test; FC > 1.2, Bonferroni corrected p < 0.05; ***p < 0.01). Adopted from (Nikitopoulou, Fanidis et al. 2021).

Despite the small numbers of dendritic cells available in the aforementioned datasets (Figure 49), we were able to identify a statistically significant *ENPP2* increase in patients' peripheral pDCs compared to those of healthy individuals. *ENPP2* expression levels did not reach statistical significance in other tissues (Figure 50A-D). To avoid the bottleneck of small cell size, we extended the searching space comparing *ENPP2* transcript count between COVID-19 and healthy individual DCs in general. *ENPP2* mRNA levels were once again found elevated in COVID-19 originating cells (Figure 50E). Strikingly, when we examined multiple immune

cell types from a big number of healthy samples (Ota, Nagafuchi et al. 2021), DCs emerged as the main *ENPP2* producer (Figure S17A). DCs also seem to express high levels of LPAR2, one of the five LPA receptors reported (Figure S17B). Apart from COVID-19, a boost in *ENPP2* expression from pDCs has also been observed in inflammatory diseases, such as systemic lupus erythematosus (SLE), adult-onset Still's disease (AOSD), mixed connective tissue disease (MCTD) and idiopathic inflammatory myopathy (IIM) when compared to steady state conditions (Figure S17C). Last, in order to identify potential *ENPP2* interacting genes we performed pairwise DEA between COVID-19 and control DCs, positive or not to *ENPP2*. We discovered that transmembrane protein 176B (TMEM176B) and CD1a, were the only up-regulated genes following the same pattern with *ENPP2* in the COVID-19 milieu (Figure 50F).

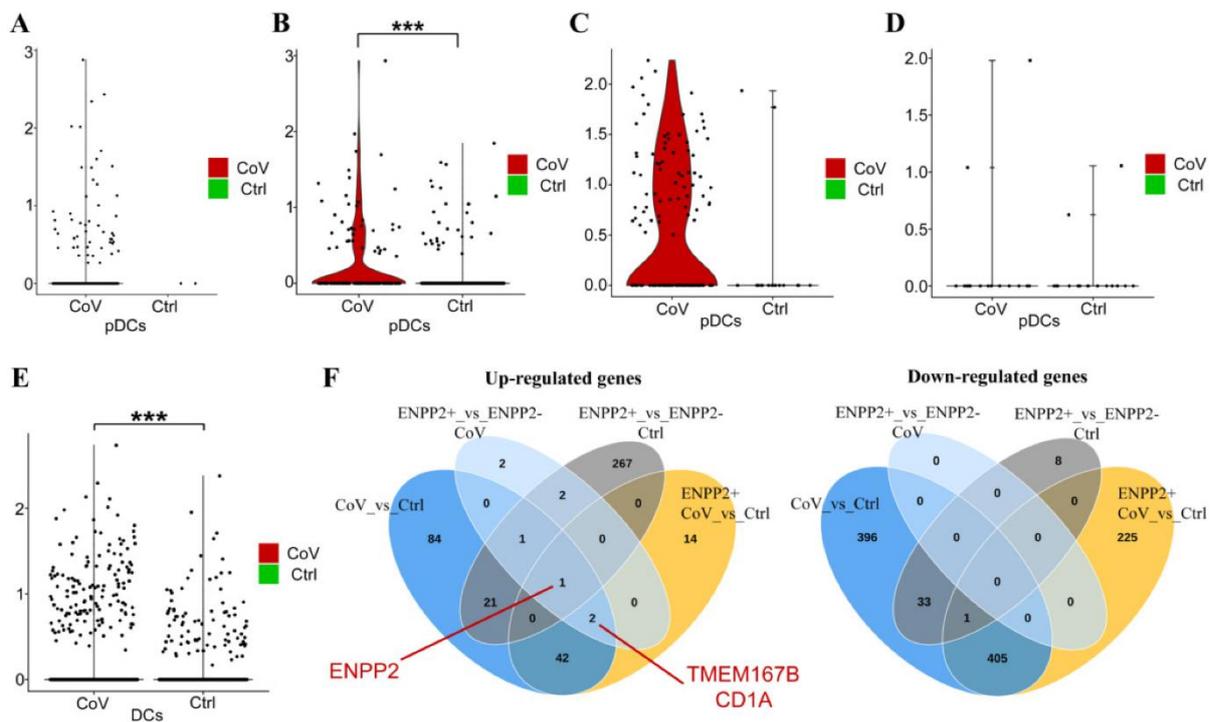


Figure 50 *ENPP2* expression is increased in dendritic cells of SARS-CoV-2 affected individuals. **A–D.** *ENPP2* differential expression in pDCs isolated from nasopharyngeal swab, PBMC, BAL fluid and lung tissue samples of COVID-19 patients versus healthy controls. *ENPP2* transcription was up-regulated in peripheral (B) patient-originated pDCs. **E.** *ENPP2* was over-expressed in pulmonary DCs of patients compared to healthy counterparts. **F.** Deregulated genes in lung DCs (Wilcoxon rank sum test; FC > 1.2; Bonferroni corrected p < 0.05). Adopted from (Nikitopoulou, Fanidis et al. 2021).

To sum up, *ENPP2* is actively transcribed by peripheral and respiratory track immune cells. From the identified *ENPP2*⁺ cell types, DCs repeatedly emerge as the most significant producer under both steady state and pathogenic conditions. More importantly, *ENPP2* transcription intensifies during SARS-CoV-2 infected individuals compared to healthy controls. Last, according to DEA results it seems that *ENPP2* affects DCs homeostasis during COVID-19, potentially affecting patients via the aforementioned cell type.

3.10 LPA and renal fibrosis

Located downstream of ATX, LPA signaling has been found to participate into various chronic diseases including rheumatoid arthritis, pulmonary and lung fibrosis (Magkrioti, Galaris et al. 2019). In parallel, accumulating evidence supports its implication in CKD (Zhang, Chen et al. 2016, Zhang, Wang et al. 2017) and the regulation of kidney TECs' secretome (Pradere, Klein

et al. 2007, Sakai, Chun et al. 2017). Motivated by these findings, we used kidney TEC's as a means of studying ATX signaling effects on signaling pathways and secreted molecules.

3.10.1 HKC-8 proteome changes induced by 176 stimulants

In two experimental setups, human renal proximal tubular epithelial cells (HKC-8) were exposed to 175 stimulants (Table S20) and LPA species 18:1 (10 μ M). In the first protocol, cells were incubated for 24 hours and 32 extracellular factors were measured in their supernatants. In the second experiment, cells were treated for 5 or 25 minutes and 27 intracellular phosphoproteins were measured using cellular lysates. Quantification of both secreted and intracellular proteins was accomplished by a custom multiplex ELISA. For data normalization purposes, per stimulus-measured abundance of each molecule was divided with same molecule levels from untreated cells (median of the control wells) (Figure 51-52). Heatmaps' color coding reflects the response intensity of each quantified molecule, with light blue and dark red corresponding to no/low and very high activation, respectively. A response ratio threshold of 1.5 gave a stable number of activations, adequate for downstream analysis (Figure S18).

According to the performed measurements, HKC-8 cells actively responded to %61 of the 176 stimuli, with 76 of the latter affecting the secreted (Figure 51) and 71 the intracellular analytes (Figure 52). More specifically, hepatocyte growth factor (HGF), interleukin 1 β , FSL1 and TNF α were the top four stimulants with 18, 17, 13 and 12 unique activations, respectively. From an analytes perspective, CXCL10 was the secreted molecule with most responses (34 the number), while JUN was the phosphoprotein with the most activations (29 the number).

To summarize, the *ad hoc* developed multiplex ELISA successfully detected treatment-induced proteome activations, while selected stimulants produced an adequate number of changes to use for downstream analyses.

3.10.2 LPA-driven changes in HKC-8 secretome and proteome

Focusing on LPA, we initially treated HKC-8 cells with LPA 18:1, enriched cells' secretome for IL-6, IL-8, CLL2, CCL3 and CXCL10 (Figure 52). Phosphorylation of JUN, I κ B α , ERK1 and CREB1 was also induced (Figure 52). Following the above responses, the cell line was exposed 16:0 or 20:4 LPA species (1.5 FC threshold for active signals) (Figure 53-54). Examination of the LPA species, revealed that secreted CLL3, IL-6 and IL-8 had an increased abundance in all three treatment regimens compared to untreated cells (Figure 53). On the other hand, secreted ICAM1 was significantly induced by 16:0 and 20:4 LPA (Figure 53). CXCL10 secretion was activated by 18:1, validating the up-stream observations.

Examining the phosphorylation patterns activated by the three LPA species, no treatment induced any change of important effect size in any of the examined time-points. However, 5 minute treatment with LPA 18:1 resulted in statistically significant up-regulation of MEK1 and CREB1 (Figure 54A). On the second time-point, statistically significant changes were only witnessed by LPA 16:0 and 20:4 treatments, as can be seen from (Figure 54B).

In conclusion, treatment with LPA 18:1 consistently affected HKC-8 cells inducing the secretion of pro-inflammatory molecules. No reproducible pattern was described for intracellular proteins phosphorylation. Similarly, more experimentations are needed to validate the observations made after treatment with the other two LPA species.

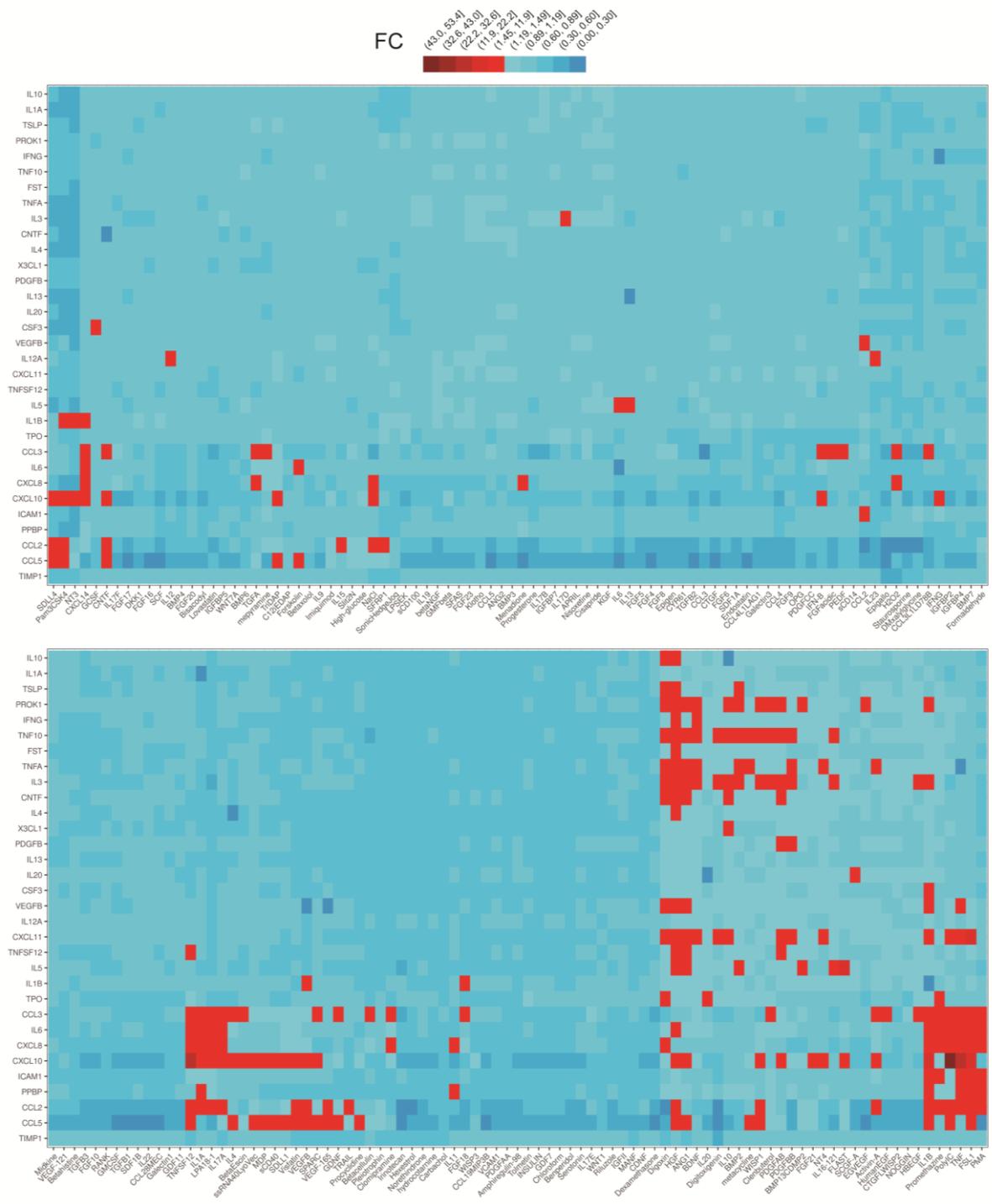


Figure 51 Impact of LPA (18:1) and 175 biological stimulants on human kidney proximal tubular epithelial cells secretome (32 factors). Red color indicates active signals (FC ≥ 1.5). Adopted from (Magkrioti, Antonopoulou et al. 2022).

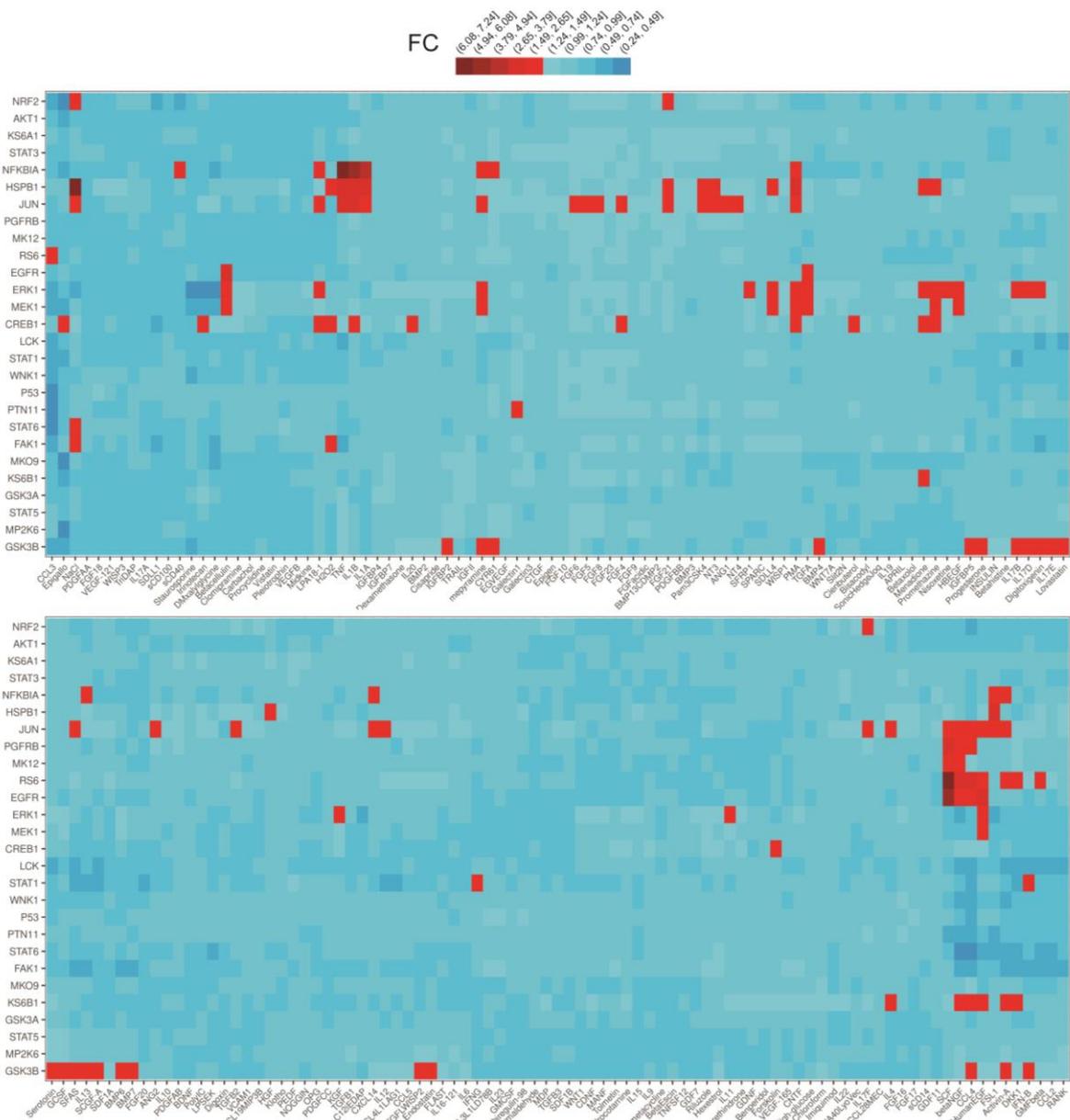


Figure 52 Impact of LPA (18:1) and 175 biological stimuli on the phosphorylation of 27 pathway hubs from proximal tubular epithelial cells. Red color indicates active signals ($FC \geq 1.5$). A signal was included if active in at least one treatment time-point Adopted from (Magkrioti, Antonopoulou et al. 2022).

3.10.3 HKC-8 responses to the other stimuli

In an attempt to group kidney responses into biologically meaningful groups, we concatenated results from both experimental setups (Figure 51-52) and isolated 108 stimuli, each one causing at least a single activation event. Subsequently, we transformed ELISA measurements into binary values representing activation (1) or impassiveness (0) and used them for stimuli unsupervised divisive clustering with the Gower's metric. As can be seen from (Figure 55), three big stimuli clusters were revealed. Interestingly, LPA clustered along with several pro-inflammatory (PMA, $IL-1\alpha$, $IL-1\beta$, $IL-17\alpha$, $TNF\alpha$ and CXCL14/BRAK) and immune system-related stimulants, such as FSL1, PolyIC, TNFSF12/TWEAK and promethazine (cluster II). Same cluster stimulants induced a consistent pattern of activated secretions ($IL-6$, $IL-8$, CCL3,

CCL2 and CXCL10). Regarding post-translational modifications, JUN and IκBα phosphorylation was induced by most stimuli, while that of CREB1 was affected by promethazine, PMA, IL-1β and LPA.

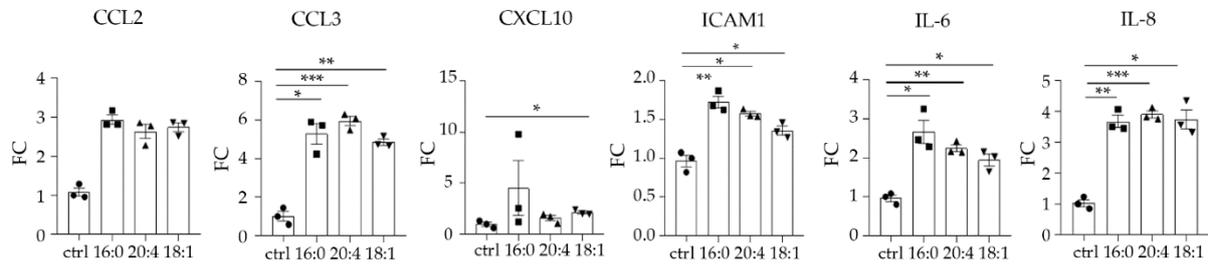


Figure 53 LPA species (16:0, 20:4, 18:1) intensify chemokine secretion from human kidney proximal tubular epithelial cells. Cell were treated with 10 μM of each LPA species for 24 hours. Normally distributed data were assessed using Brown–Forsythe’s and Welch’s ANOVA followed by Dunnett’s post hoc test, while non-normally distributed data with Kruskal–Wallis test (*p < 0.05, **p < 0.01, ***p < 0.001). Circles, squares, upward triangles and downward triangles iconize control, LPA 16:0, LPA 20:4 and LPA 18:1 values, respectively. Adopted from (Magkrioti, Antonopoulou et al. 2022).

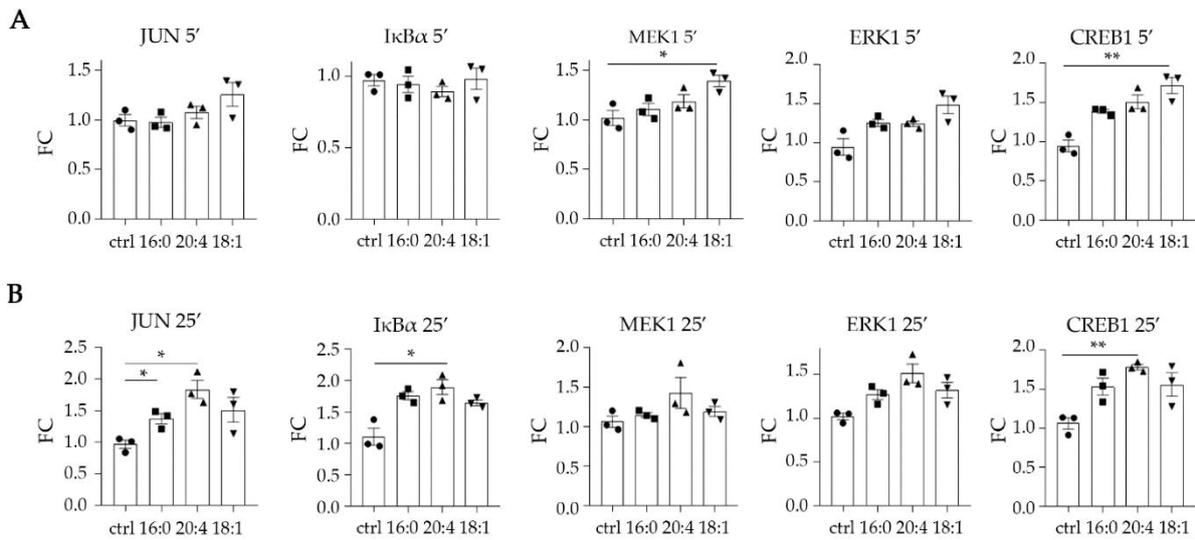


Figure 54 LPA species (16:0, 20:4, 18:1) induce the phosphorylation of transcription factors and signaling kinases of human kidney proximal tubular epithelial cells. Cell were treated with 10 μM of each LPA species for 5 (A) or 25 min (B), and the phosphorylation was assessed with multiplex ELISA in triplicates. Normally distributed data were assessed using Brown–Forsythe’s and Welch’s ANOVA followed by Dunnett’s post hoc test, while non-normally distributed data with Kruskal–Wallis test (*p < 0.05, **p < 0.01). Circles, squares, upward triangles and downward triangles iconize control, LPA 16:0, LPA 20:4 and LPA 18:1 values, respectively. Adopted from (Magkrioti, Antonopoulou et al. 2022).

Similarly to the 2nd one, cluster I consisted of a rather small group of stimulants (ANG1, HGF, PDGF-AB, PDGF-BB, digoxin, BMP2, BDNF and IL-10). Importantly, multiple activations (Figure 56) were induced by ANG1 and BMP2, that have been shown to have a protective role in models of renal injury (He, Zhang et al. 2019) and to polarize adult renal progenitor cells (Simone, Cosola et al. 2012), respectively.

Last, cluster III – the biggest of the three – included multiple protein activation events, many of which are directly or indirectly linked to kidney fibrosis. Indicatively, WISP1 that has been reported to regulate kidney fibrosis (Zhong, Tu et al. 2017) induced the secretion of CCL2, CCL5, CXCL10, TNF10, PROK1 and IL-3. Moreover, activin A which induced both phosphorylation (GSK3B, KS6B1, RS6, JUN, IκBα) and secretion (CCL3, CXCL10, CCL2) of multiple molecules has been observed to activate renal fibroblasts during kidney fibrosis (Yamashita, Maeshima et al. 2004).

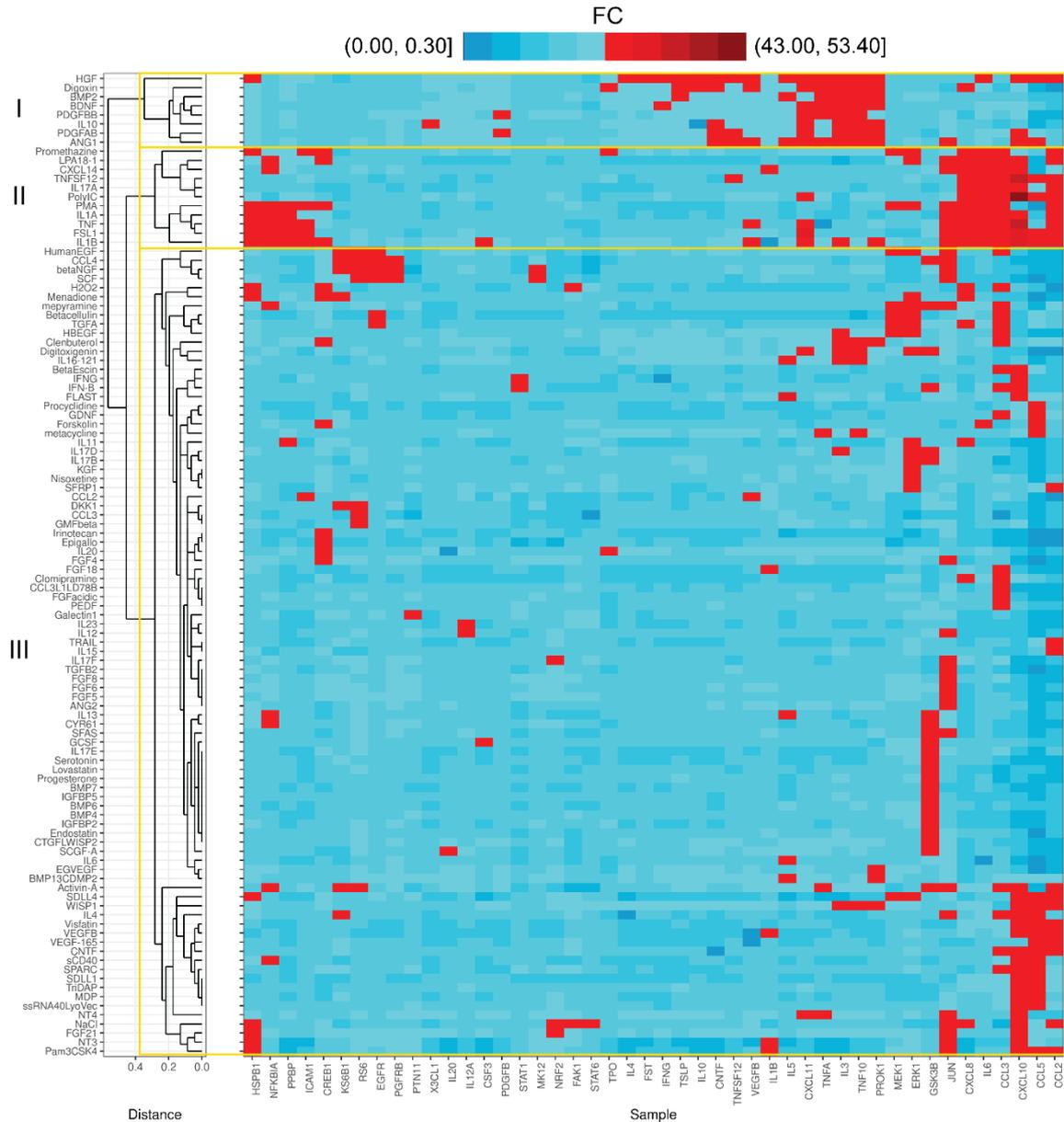


Figure 55 LPA affects human kidney proximal tubular epithelial cells in a fashion similar to that of pro-inflammatory stimuli. Active stimuli are clustered in three groups defined by latin enumeration. In active signals/stimuli are not included. Pairwise stimuli distance was calculated on binary transformed fold change values using Gower’s metric prior to divisive clustering. Adopted from (Magkrioti, Antonopoulou et al. 2022).

3.10.4 Focus on cluster II stimulants

Subsequently, we attempted to dissect the effects of LPA and similar signaling (cluster II) in the kidneys. Initially, we performed PA on the active responses of cluster II genomic stimulants with GO and KEGG as reference databases (Figure S19A). By examining the most highly enriched pathways/ontologies, we noticed that cluster II was most probably related to inflammatory and pathogen invasion-related processes (Figure S19B-C). Then, we re-analyzed two microarray datasets containing CKD, focal segmental glomerulosclerosis and immunoglobulin-A nephropathy samples along with their control counterparts. Four molecules affected by cluster II stimulants (TNFSF12, IL-1 β , ICAM1, CCL5) were among the differentially expressed genes identified between pathology and control samples (Figure S20). ICAM1 was also activated by LPA proposing the existence of a common mechanism between kidney diseases and LPA signaling. Cell specificity of cluster II features was assessed by re-analysis of single cells from kidneys of CKD and control cases (Kuppe, Ibrahim et al. 2021). In brief, *CCL2*, *CCL3*, *CCL5*, *CXCL8* (coding for IL-8) and *IL-1 β* marked CD10- proximal tubule epithelial cells (PT-negative) and *VEGFB* CD10+ cells (PT-positive) (Figure S21A), validating that kidney cells can indeed produce those molecules.

Last, we performed a ligand-receptor analysis in search of underlying cell-to-cell communication networks, potentially reflecting in culture discovered stimulations. For this purpose, SingleCellSignalR database was supplemented with CellTalkDB-sourced interactions. With the exception of PT-positive and PT-negative cells all other cell type groups were clustered per population for purposes of facilitating results' interpretation. Examination of PT-marking cytokines revealed a rather universal communication network proposing PT cells as a potential starting point for both autocrine and paracrine signals towards all other cell populations (Figure S21B).

In brief, various *in silico* analyses suggested that cluster II stimuli affect immune system-related processes which potentially underly various pathological conditions. Furthermore, single cell data examination validated the capability of proximal tubular epithelial cells to secrete molecules activated by cluster II stimulants.

3.10.5 *In vitro* validation and dissection of LPA effects

With the aim of verifying the above mentioned effects, transcription of six genes was evaluated post to 1, 4, 12 and 24 h stimulation of HKC-8 cells with LPA 18:1. As can be seen in (Figure 56A), the abundance of all secreted molecules peaked after 4 hours of treatment. Measuring mRNA levels after 4h treatment with various LPA concentrations (2.5, 5 and 10 μ M) revealed that 10 μ M of LPA cause maximum transcriptional activation of almost all examined genes (Figure 56B). Finally, HKC-8 cell surface LPA receptors were quantified using the same RNA samples. LPAR2 and LPAR6 were found as the primary LPA receptors, yet no significant change was detected post to LPA treatment (Figure S22).

To intra-cellularly dissect LPA signaling we measured mRNA levels of previously identified LPA-affected phosphoproteins post to LPA, LPA/Ki16425 and LPA/H2L5186303 treatment. Ki16425 and H2L5186303 were selected as inhibitors of LPAR1/3 and LPAR2, respectively. LPAR1/3 were responsible for the regulation of *CCL3*, *CXCL8* and *IL-6*, with LPAR2 acting as an extra switch that controlled *IL-6* transcription (Figure S23). Participation of other LPARs in the LPA signaling cascade cannot be excluded due to lack of proper inhibitors.

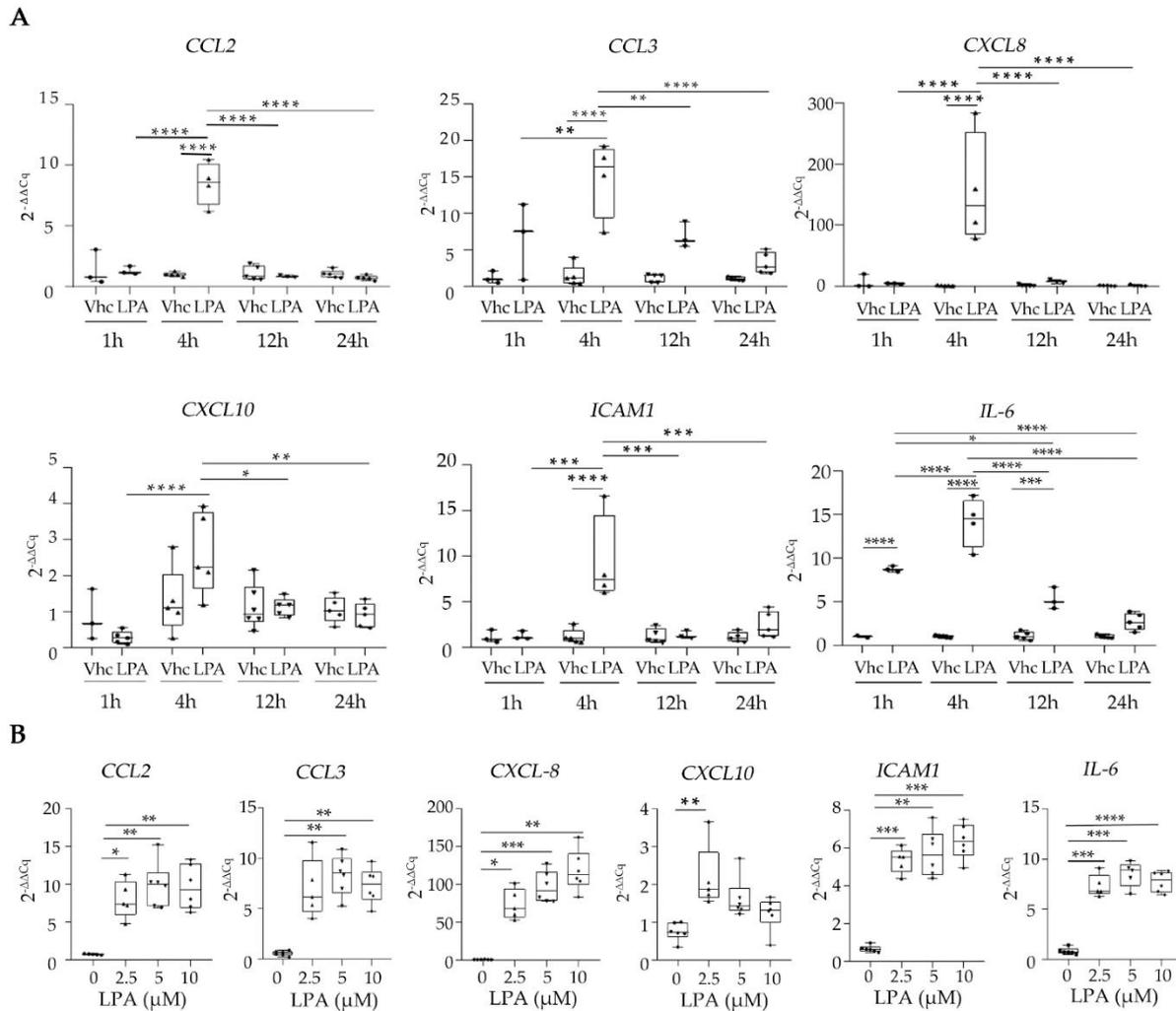


Figure 56 Chemokines production from human kidney proximal tubular epithelial cells is affected by various LPA signaling parameters. A. Cells were treated with 10 μ M of LPA for 1, 4, 12, and 24 h. **B.** Cells were treated with 2.5, 5, and 10 μ M LPA for 4 h. In both (A-B) control cells were intubated with chloroform. Gene expression was measured with RT-qPCR and Cq values were normalized against those of *B2M*. In (A) circles, upward triangles, downward triangles and diamonds iconize 1, 4, 12 and 24 hours of incubation with LPA, respectively. In (B) circles, upward triangles, downward triangles and diamonds represent treatment with 0, 2.5, 5 and 10 μ M of LPA, respectively. Statistical significance was assessed with 2-way ANOVA and Tukey's post hoc test (A) and with Brown-Forsythe's and Welch's test or the Kruskal–Wallis test depending on data normality (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Adopted from (Magkrioti, Antonopoulou et al. 2022).

Last, to validate changes in protein phosphorylation, we treated HKC-8 cells with SP600125 (a JNK inhibitor), PD98059 (a MEK/ERK inhibitor), JSH23 (an NF κ B inhibitor) or 666-15 (a Creb inhibitor). PD98059, JSH23 and 666-15 treatments lessened *CCL2*, *CCL3*, *CXCL8* and *ICAMI* expression (Figure 57A-C). In parallel, *CCL3* and *ICAMI* transcription was obstructed by SP600125 (Figure 57D). These results propose that the above targeted pathways, JNK/c-JUN, MEK/ERK, NF κ B and CREB act downstream of LPA and regulate cellular secretion. Indeed, a literature search indicated that *CCL2*, *CCL3*, *CXCL8* and *ICAMI* are co-regulated by the above mentioned pathways (Figure 58).

In conclusion, we were able to validate LPA 18:1 effects on multiple molecules at both the transcriptional and translational level. A 4h treatment period with 10 μM of LPA was required for HKC-8 cells' transcription to be maximally affected, but without any significant change in LPA receptors produced mRNAs. Use of pathway inhibitors further supported our findings by achieving the exact opposite results from the treatment assays.

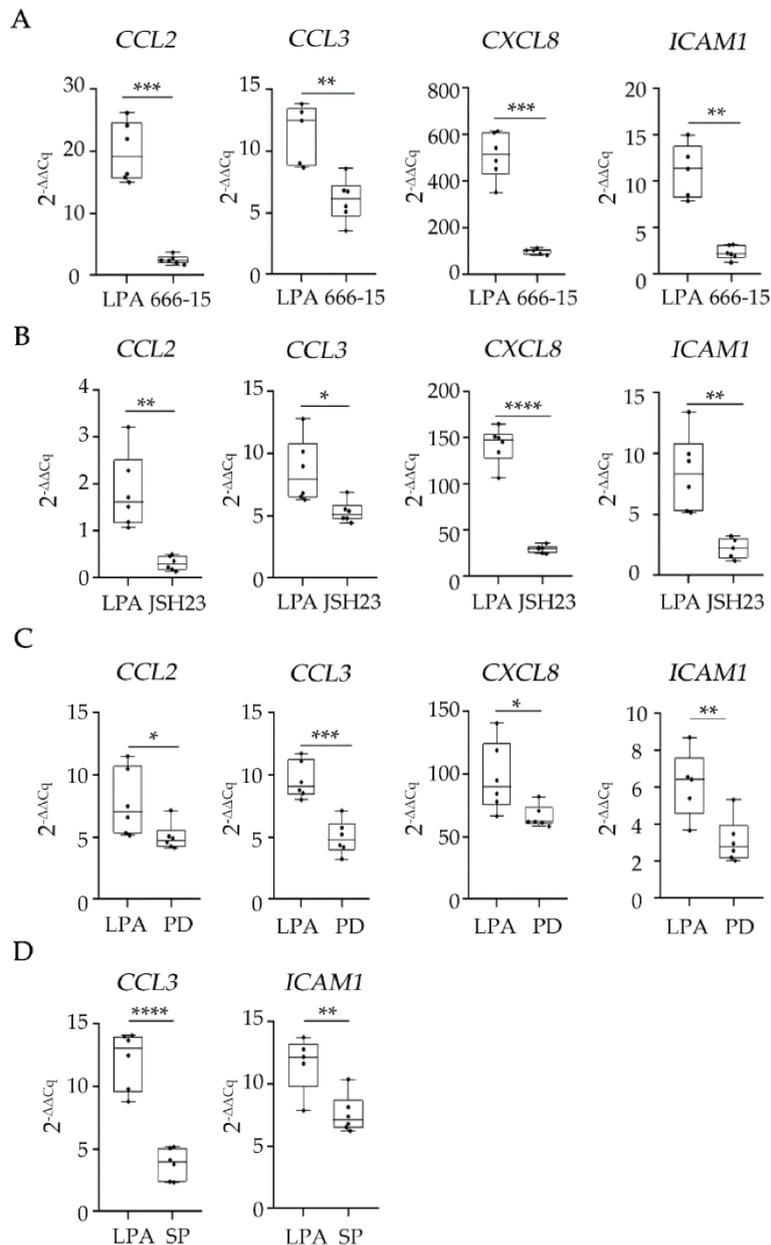


Figure 57 Obstruction of LPA-stimulated intracellular signaling pathways in kidney proximal tubular epithelial cells. A-D. Cells were pretreated for 1 h with 10 μM of 666-15 (CREB1 inhibitor), 100 μM of JSH23 (NF κ B inhibitor), 50 μM of PD98059 (MEK/ERK inhibitor) and 50 μM of SP600125 (JNK inhibitor), respectively. Subsequently they were treated for 4 h with 10 μM of LPA (final concentration). Gene expression was measured with RT-qPCR, while per gene Cq values were normalized against those of *B2M*. In cases of normality, statistical analysis was performed with unpaired t-test or Welch's test; Mann-Whitney test was performed in all other cases (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Adopted from (Magkrioti, Antonopoulou et al. 2022).

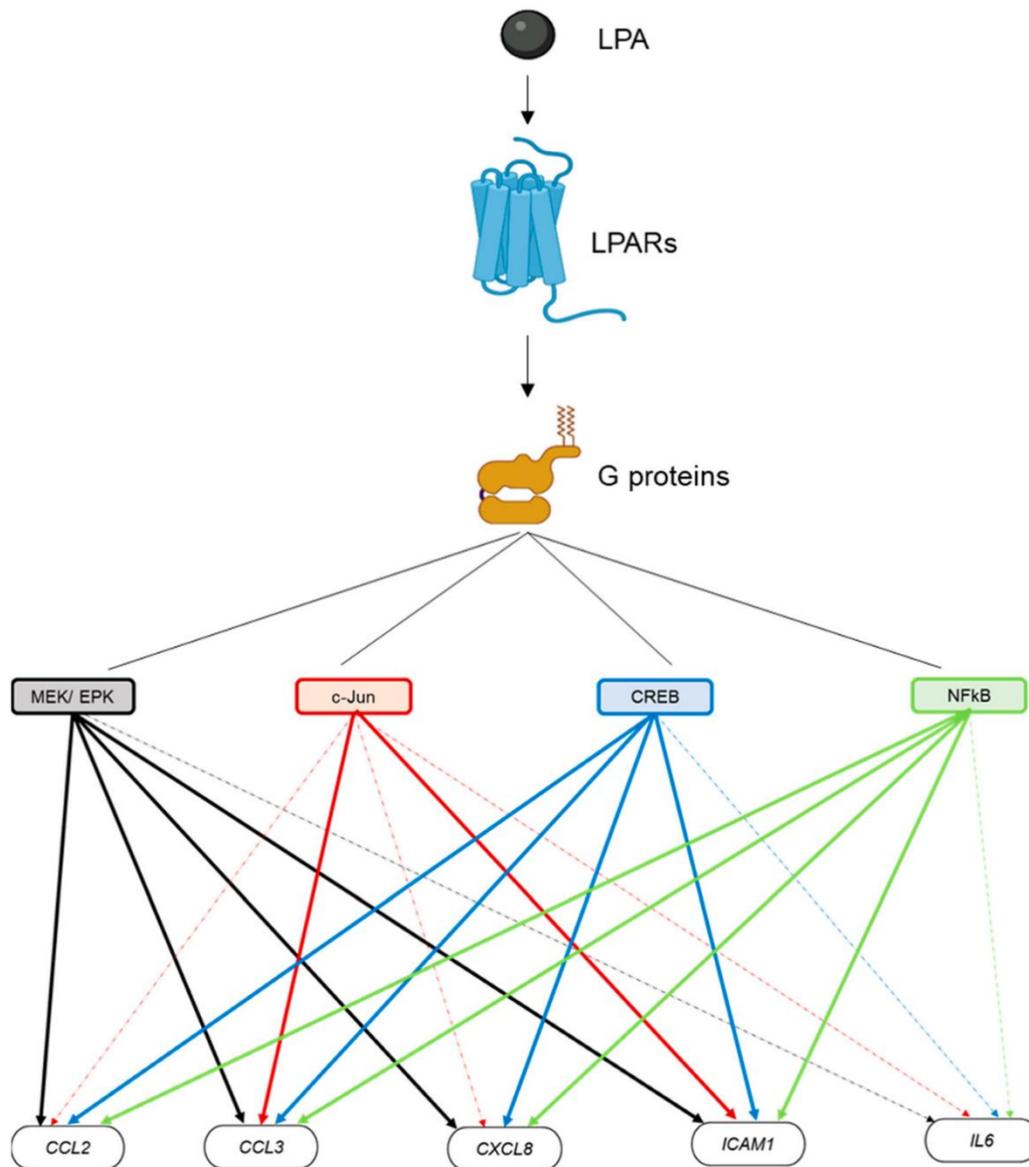


Figure 58 Visualization of LPA-stimulated signaling pathways in human kidney proximal tubular epithelial cells. Adopted from (Magkrioti, Antonopoulou et al. 2022).

4. Discussion

4.1 New IPF targets from public omics resources

IPF is a lethal progressive fibrotic disease of unknown cause, with a heavily heterogeneous between patients progress (Podolanczuk, Thomson et al. 2023). Although an orphan disease it bears a significant socioeconomic impact (Raimundo, Chang et al. 2016, Hilberg, Bendstrup et al. 2018), especially for aging societies, while its worldwide mortality rate has been reported to steadily increase (Hutchinson, McKeever et al. 2014). As a curative treatment has yet to be found, medical doctors can only subscribe anti-fibrotic agents, such as nintedanib and pirfenidone, in an attempt to delay disease progression in exchange for many times severe adverse effects (Podolanczuk, Thomson et al. 2023).

Currently, multiple genomic targets are known to be implicated in the establishment and progression of PF (Distler, Györfi et al. 2019, Henderson, Rieder et al. 2020, Bowman, Echt et al. 2021). A great percentage of these gene markers has been identified via omics technologies;

primarily transcriptomics (Vukmirovic and Kaminski 2018) and secondarily proteomics (Khan, Dasgupta et al. 2021). One of the elementary corollaries of such an active research was the accumulation of large data volumes which had not been extensively mined for latent yet significant information.

As the scientific community begun to understand the additive value of these publicly available resources, attempts were made to combine and interpret them. The purpose of most such research deliverables was to discover biomarkers that could potentially unlock new treatment options and even explain IPF's driving force. Based on the importance of highly interconnected genes (aka hub genes) (Carlson, Zhang et al. 2006, Helsen, Frickel et al. 2019), several of these transcriptomic-oriented publications exploited just a handful of datasets in order to perform DEA, create interaction networks, identify and report their hub genes (Li, Liu et al. 2020, Xu, Mo et al. 2020, Cui, Ji et al. 2021, Zhu, Xu et al. 2021). Others such as (McDonough, Kaminski et al. 2019, Wan, Huang et al. 2021) have technically enriched their analysis with the use of WGCNA, while (Karatzas, Bourdakou et al. 2017) exploited transcriptomic data from multiple publications aiming at drug repurposing.

The aforementioned attempts may have provided several clues about transcriptome organization and regulation during IPF, yet they are highly incomprehensive as they focus on specific data collections. To provide a more holistic view of PF, (Villaseñor-Altamirano, Moretto et al. 2020) collected “without any restriction” human COPD and IPF transcriptomic datasets from GEO. These data are accessible via PulmonDB, a web-tool that supports data mining and visualization powered by the Clustergrammer package (Fernandez, Gundersen et al. 2017). Nevertheless, even PulmonDB does not address all PF-related omics datasets as it excludes animal model, proteomics and scRNA-seq datasets. At the same time, its online exploratory capabilities are limited and acquisition of the hosted pre-processed data requires a certain level of computational skills.

In this project, we planned on creating a comprehensive collection of PF datasets and utilize it for the proposal of new biomarker genes. Working towards that goal, we initially developed Fibromine, a database and accompanying toolkit dedicated to the collection, integration and presentation of IPF omics data. For its creation we have manually scrutinized the scientific literature for related publications and carefully selected the best publicly-available transcriptomic and proteomic datasets. Transcriptomic data were consistently re-analyzed and quality controlled, both manually and algorithmically, while the far fewer proteomic collections were handled separately. Outlier samples were removed from all transcriptomic datasets. Data collections of low quality revealed during re-analysis were filtered out as well. For the harmonization of the supported conditions defined via retrieved meta-data (clinical phenotypes, experimental conditions) we have created a structured vocabulary of terms which was also used for the description of the presented DEA/differential abundance comparisons. Collectively, our database hosts 47 human/human-origin and 13 mouse collections spanning 42 distinct condition comparisons and 5 omics technologies.

For proper data presentation and handling, we have armed Fibromine with a great variety of dataset-specific and general-purpose metadata. The latter were either manually gleaned from the respective publications or systematically collected from publicly-accessible databases using *ad hoc* created scripts. Ensembl (Yates, Achuthan et al. 2020), STRING (Szklarczyk,

Gable et al. 2019), miRBase (Kozomara, Birgaoanu et al. 2019) and RefSeq (O'Leary, Wright et al. 2016) are some of the sources utilized.

To enable on-the-fly mining and integration of Fibromine data, we crafted the homonym web-based Shiny application. Structured around three main explorers, the application offers, among others, identification of cDEGs and cDEPs, exploration of gene-specific expression patterns on a bulk and single cell level and the creation of disease-specific PPI networks. Moreover, the user can explore deregulated miRNA-mRNA pairs, along with WGCNA of human and mouse origin. DEA data from three integrated scRNA-seq datasets are also easily accessible. Some of the rest available services are the real-time PA supported by *enrichR* (Chen, Tan et al. 2013), as well as the creation of dataset-specific volcano and heatmap plots for bulk transcriptomic datasets. All pre-processed bulk data can be downloaded and used for any downstream application.

Compared to its predecessors, Fibromine presents with multiple advantages and extra features. Initially, it is the first comprehensive collection of high-quality, manually curated PF omic datasets. Thanks to the inclusion of multiple experimental models and platforms, the user can directly compare their findings in order to validate wet laboratory observations and/or form new hypotheses. In addition, apart from the multifaceted data exploration tools provided, the user can integrate differential expression/abundance data from any hosted datasets irrespective of the featured species, platform or technology. This is achieved by the calculation of $FC_{\text{consensus}}$, a metric that cannot be affected by biases potentially introduced through batch or technology particularities. Owing to this data combination system, Fibromine is a highly flexible and extendable toolkit capable of hosting a great variety of data without the need of reanalyzing its existing database. To the best of our knowledge, Fibromine offers the first condition-specific PF-oriented PPI network creation tool, enabling the projection of condensed knowledge to the protein interaction space. Moreover, as part of Fibromine, we were the first to attempt a transcriptomic datasets benchmarking. Based on this qualitative system future users can quickly identify similar datasets to use. Last, Fibromine was designed to be as transparent and user-tailored as possible, by providing backend algorithmic details and enabling dynamic modification of the vast majority of its parameters, respectively.

Despite its many advantages, Fibromine is also characterized by certain drawbacks/ limitations. As far as the animal model datasets used are concerned, exact condition matching cannot be guaranteed as various bleomycin treatment protocols have been utilized. Even in cases of same BLM quantities usage, track of administration may differ thus potentially introducing differences to the examined phenotype. Moreover, Fibromine PPI networks annotation is based on integrated values of differential gene expression and not on their proteomic counterparts. While this pipeline just reflects what the proteomic landscape could be, usage of transcriptomic data was unfortunately the best possible option given the limited number of unique proteins (693) assessed by an even smaller number of proteomic datasets. In addition, while Fibromine hosts a great proportion of the high quality PF omics datasets that are publicly available, there are certain biological modalities that are not adequately represented. Concerning scRNA-seq, the small number of published datasets during Fibromine's creation was the inclusion-limiting factor.

It is important to note that Fibromine's operational processes are subject to certain assumptions. For example, murine datasets benchmarking hypothesize that mouse pro-fibrotic features exert

an effect analogous to that of their human homologues. Furthermore, cDEG identification exploit nominal instead of multiple testing corrected p-values, so as to avoid usage of a hard threshold that could potentially exclude marginally significant observations. While consultation of FDR corrected p-values presented at *Transcriptomics/Proteomics analytically* tabs is highly recommended, cDEGs identification is simply based on the verifying value of an observation's independent repetition.

As described above Fibromine can quickly and painstakingly provide any user with a list of the most promising IPF molecular elements from a huge pool of available options. Nevertheless, application's use has shown that the produced cDEG/cDEP collections can still some times be of a size that renders targets selection a rather subjective process. To address this limitation, we decided to attempt target discovery with the assistance of a ML algorithm. During the past years, ML strategies have successfully been used in PF research with most applications using image data (Walsh, Calandriello et al. 2018, Huang, Lee et al. 2020, Barnes, Humphries et al. 2023, Pan, Hofmanninger et al. 2023). Endotyping (Kim, Herazo-Maya et al. 2015) and differential diagnosis (Kim, Diggans et al. 2015) predicted from gene expression values are other less frequent ML objectives in the IPF research field.

Thus, in continuance to Fibromine we have trained an XGBoost ensemble learner to classify IPF and control samples. Supervised learning was based on a subset of IPF_vs_Ctrl Fibromine cDEGs identified from a benchmarking-proposed set of transcriptomic data collections. To form the final training feature set, cDEGs list underwent a second round of selection according to an *ad hoc* formulated integrated score (S_i) that combined knowledge between differential expression and semantics similarity. Supervised training and testing were succeeded by an external models' validation which utilized an independent cohort of samples. Subsequently, SHAP values were calculated to interpret models' very good performance yielding ten marker sets which were then integrated by two top and a custom developed ranking aggregation methods. According to functional analyses, both well established and less known DEGs lied among the prioritized features. In addition, we have selected from the top marker genes those shared between ranking aggregation methods and used them to repeat XGBoost training/testing and validation process. Performance of the new model was compared to that of others which were instructed to discern phenotypes from the expression of biomarker sets computationally-shaped from past publications. Our algorithm was among the best performers in the internal testing set with distinguished performance during external validation. Last, the expression of SHAP value-prioritized marker genes was correlated to lung function measurements revealing several important associations.

Apart from producing two novel sets of IPF biomarkers, the aforementioned work revealed that ML algorithms and related methods can be used for phenotype classification when powered by gene expression data. In this context, we have validated the previously reported successful use of XGBoost with transcriptomics data (Dimopoulos, Koukoutegos et al. 2021). As far as we are able to know, we are also the first to use SHAP values for disease target selection, potentially setting the foundations for their broader usage. Moreover, the small performance differences witnessed between models is in agreement with observations from already published ML applications (Pezoulas, Papaloukas et al. 2021). Last and according to their performance, ranking aggregation methods seem to be a useful and straight-forward computational tool for the integration of DEG lists originating from distinct data sources.

Supplementary to the above observations and suggestions and in favor of any similar future attempts, it is important to note some ML technicalities encountered during the herein discussed application. To begin with, ML algorithms must be carefully selected in order to be able to address DEG characteristics, such as their inherent co-linearity. XGBoost was purposefully distinguished due to its robust nature against overfitting and multi-co-linearity phenomena (Parsa, Movahedi et al. 2020). Furthermore, the number of training features must be carefully selected to achieve a good model performance, yet without restricting the pool of candidate biomarkers.

As far as the biological component of this research attempt is concerned, we were able to suggest a long and a short/core set of target genes consisting of 76 and 13 features, respectively. Interestingly, both over- and under-expressed genes are hosted in both groups potentially reflecting simultaneous differential regulation of anti- and pro-fibrotic genes during PF. Limited intersection of the 76 features with those mined from PubMed abstract corpus suggests the identification of latent disease targets that require future experimental assessment. These findings should be interpreted with caution as an abstract corpus has a finite capacity of recording all disease-related transcriptomic elements that have been found. Nevertheless, there are certain promising targets, such as *IL13RA2*, *DCXR* and *SFRP2*, that are directly or indirectly connected to fibrosis and have not yet been extensively researched. In detail, *IL13RA2* that was here significantly correlated with lung function metrics has been proposed to play a pro-fibrotic role in the colon (Fichtner-Feigl, Strober et al. 2006), while *DCXR* suppression has been reported to ameliorate kidney fibrosis (Odani, Asami et al. 2008). Among the top-prioritized features lies *SFRP2*, a gene that is well established as a marker of myofibroblast progenitors in systemic skin sclerosis (Tabib, Huang et al. 2021) and has been also detected in lung fibroblasts (Mayr, Simon et al. 2021), thus potentially participating in fibrosis establishment. Last, there are under-researched targets even between those detected during text mining. Indicatively, *CTHRC1* is a high-ranked feature that although over-expressed during IPF was only recently shown to mark pro-fibrotic genes in the fibrotic lung (Tsukui, Sun et al. 2020).

Conclusively, we have developed Fibromine, an open-access database of manually curated and consistently re-analyzed omics datasets revolving around PF. These data can be accessed via various data integration and exploration tools provided by the homonym web-based toolkit. More importantly, we have showcased that Fibromine-integrated data can be creatively exploited for IPF research, such as the training of a ML algorithm for phenotype classification and biomarker discovery. Future research endeavors needs to be done in order to address limitations of the herein presented design, as well as for identified disease targets wet laboratory validation.

4.2 Methylation-dependent transcriptional control of *ENPP2* during cancer

DNA methylation is a significant tier of transcription regulation both during steady state and pathological conditions. Despite its undisputable importance for human cells, there are still multiple open questions regarding the functions of DNA methylation (Schübeler 2015). It is now known that DNA methylation patterns do change in lung fibroblasts during IPF (Huan, Yang et al. 2015, Lee, Son et al. 2019, Negreros, Hagoood et al. 2019), while methylation changes observed in the fibrotic lung are sometimes comparable to those of tumor cells (Rabinovich, Kapetanaki et al. 2012).

ENPP2 encodes for ATX, a glycoprotein responsible for the regulation of various processes in the human organism. Patients suffering from chronic inflammatory diseases, including IPF, have increased levels of both ATX and its effector product, LPA (Magkrioti, Galaris et al. 2019), while ATX-LPA axis is also implicated in tumor formation (Magkrioti, Oikonomou et al. 2018). The exact *ENPP2* transcriptional regulation is still under active research with previous findings pinpointing the gene as a genomic locus prone to epigenetic modifications (Parris, Kovács et al. 2014).

Motivated by the aforementioned data we examined the functional relationship of *ENPP2* methylation and expression during cancer and steady state. *In silico* comparison of publicly available healthy and tumor samples from various cancer types revealed aberrant methylation patterns during pathology. In detail, analysis of GEO-retrieved samples showed that *ENPP2* methylation followed a consistent pattern across control sampling sites, with increased and decreased levels of methyl groups in the gene body and TSS/1st exon CGs, respectively. On the other hand, and although *ENPP2* methylation during cancer seems to adhere to a malignancy-specific motif, in general, TSS/1st exon and gene body sites were hypo- and hyper-methylated, respectively, compared to healthy samples. The totality of the TSS/1st exon-located DMCs were characterized by increased methylation across cancer types such as HCC, LC and PC, in accordance to previous findings (Parris, Kovács et al. 2014, Liu, Peng et al. 2020, Nema, Shrivastava et al. 2021, Wang, Liao et al. 2021).

Subsequently, differential gene expression and differential methylation analysis of TCGA data validated that TSS and 1st exon CGs were hyper-methylated in PC, LC, HCC and BrCa samples related to their control counterparts. The opposite pattern was witnessed for *ENPP2* gene body CGs, while *ENPP2* transcription was down-regulated in PC and LC samples compared to steady state. Expression of the *ENPP2* gene followed the same expression pattern in BrCa. In addition, our results revealed a significant correlation between the hyper-methylated PA/1st exon areas and the decreased *ENPP2* transcription during LC, PC and BrCa, extending previous observations (Parris, Kovács et al. 2014, Nema, Shrivastava et al. 2021). In HCC samples, ATX expression was up-regulated irrespective of TSS and 1st exon methylation. Similar *ENPP2* expression changes have been previously reported for HCC (Wu, Xu et al. 2010). It should be noted that according to the existing bibliography, mRNA and protein levels are not straight-forward related during cancer (Chen, Gharib et al. 2002, Maier, Güell et al. 2009, Lemée, Clavreul et al. 2018, Arcos-Montoya, Wegman-Ostrosky et al. 2021). Interestingly, cg02156680 methylation obeyed to the same pattern in the BrCa, PC and HCC examined cases. Similar findings arose from the analysis of UALCAN data, as *ENPP2* was found under-expressed and over-methylated in LC and PC cases. Nevertheless, no such regulatory link was revealed from the examination of HCC samples, where both processes of DNA transcription and methylation were up-regulated. Thus, it can be suggested that *ENPP2* methylation functions in a cancer-type-specific manner.

As far as transcriptomic isoform and DNA methylation levels are concerned, our study proposed that methylation of promoter-related CGs is in general negatively related with *ENPP2* isoform expression, which could in turn be characterized as cancer-specific. In more detail, cg06998282 was the only CG whose methylation was negatively related to *ENPP2* mRNA levels in all PC, LC and HCC malignancies. Importantly, the cg02709432-centered negative correlation between the two modalities in both PC and LC, could be attributed to TF binding inhibition and thus *ENPP2* transcription (Tate and Bird 1993, Farina, Cappabianca et al. 2012).

Such obstruction of gene transcription by promoter methylation that hinders TF binding is a known regulatory mechanism (Héberlé and Bardet 2019). In our study we have *in silico* predicted 39 TFs potentially targeting *ENPP2*, with several of them known to participate into the aforementioned regulatory process (Tate and Bird 1993, Chen and O'Connor 2005).

With the aim of assessing the importance of *ENPP2* methylation for tumor prognosis, we analyzed clinical characteristics of the examined human samples. Indicatively, PC *ENPP2* methylation levels were positively related to tumor size and response to therapy, while DNA methylation was associated with advanced LC stages. In HCC, *ENPP2* methylation was in turn correlated with tumor micro-invasion. Examination of cancer cell lines with varying aggressiveness revealed heavier *ENPP2* methylation in the more invasive PC and HCC cell cultures. In a similar fashion, two TSS and a single 1st exon CG were related to BrCa metastasis. Significantly, via differential methylation analysis we suggest that hyper-methylation of *ENPP2* promoter CG sites takes place during the early and persists during latter BrCa stages, in accordance to previous studies (Ivan, Patricia et al. 2021).

In an attempt to evaluate DNA methylation drugability, we showed that *ENPP2* de-methylation can be achieved through 5-AZA treatment of three colon cancer cell lines. Starting from this observation and by taking into consideration the pathologic role of ATX in several contexts (Nikitopoulou, Oikonomou et al. 2012, Oikonomou, Mouratis et al. 2012, Magkrioti, Oikonomou et al. 2018), future studies should examine *ENPP2* de-methylation as a novel cancer therapeutic option.

Last, based on a previous report of consistent methylation between ccfDNA and affected tissue DNA (Panagopoulou, Karaglani et al. 2019), we re-analyzed a publicly available dataset of ccfDNA samples and successfully detected heavier methylation of promoter-related CGs during BrCa compared to steady state. To replicate *in vitro* the above *in silico* findings, we used a qMSP assay to assess the methylation status of a 1st exon DMC, cg02534163. Hyper-methylation was witnessed more often in BrCa affected than in cancer-free individuals in contrast to an older study (Wang, Liao et al. 2021). This discrepancy can be attributed to population and even methodological differences of the two papers. In addition, increased methylation was found in neoadjuvant and metastatic samples compared to adjuvant and control cases.

Ultimately, we have shown that *ENPP2* methylation changes drastically during cancer, a change that seems to follow a malignancy-specific pattern. Increased methylation of TSS/1st exon CGs related with decreased expression in PC and LC cancer samples, a relationship that was not validated in HCC. The above relationship could potentially be manifested, at least in some cases, by the inhibition of transcription-necessary TFs. Significantly, we propose that *ENPP2* methylation can be used as a cancer prognostic biomarker, motivating for more related studies. Future larger studies are needed in order to evaluate the clinical importance of *ENPP2* methylation in BrCa and its measurement in ccfDNA as an easily accessible biomarker.

4.3 The lung microbiome

The microbiome is a “system” so important for the homeostasis of the human organism that disturbance of the fragile balance between its microbe-members has been related to multiple human diseases (Cho and Blaser 2012, Fan and Pedersen 2021). Different organs, such as the gut, the lungs and the liver not only host their own specific microbiome, but they also exploit it for communication purposes (Tripathi, Debelius et al. 2018). In the IPF context, a

longitudinal study witnessed host response to alternations of BAL microbes and even proposed lower-airways microbiome to cause a continuous lung injury during pathology (Molyneaux, Willis-Owen et al. 2017). Lung microbiome *per se* is characterized by decreased abundance during IPF (Tong, Su et al. 2019).

In the current thesis, we have attempted to examine in parallel the microbiome of the gut, lungs and liver in search of an association between them, as well as a functional/causal relationship with obesity and fibrosis. Towards that goal, we have collected samples from HFD-driven obese and CD-fed control mice. These two phenotypes were selected as obesity is known to affect the microbiome (Stanislowski, Dabelea et al. 2019) and has also been connected to cystic (Litvin, Yoon et al. 2019) and lung fibrosis (Guo, Sunil et al. 2022), potentially as a chronic inflammation-causing factor. 16S rRNA sequencing was used for species detection and focus was given on seven hot spots for the achievement of better sensitivity (Gloor, Macklaim et al. 2017, Johnson, Spakowicz et al. 2019). DADA2-pre-processed 16S rRNA data were merged per diet and tissue of origin and assigned to ASVs before subjected to GCN correction. Although merged samples offered a “cleaner” picture of the examined milieu, they unfortunately limited the applicable statistical operations and thus the final inferences of the study. ASVs were preferred over operational taxonomic units, given that the latter come with various limitations, such as polarization of microbial diversity metrics (Callahan, McMurdie et al. 2017) and inadequate taxa representation (Větrovský and Baldrian 2013).

Microbiomes were then examined for their richness and diversity, with richer communities seemingly detected under HFD. This observation was reinforced by the calculated Shannon’s index only in lung tissue samples. Next, β -diversity suggested that HFD polarizes lung microbiome towards a liver-like composition. Intersection of phyla sampled from the various animal phenotypes and anatomical locations indicated that HFD increase the absolute numbers of common taxa across tissues, a phenomenon replicated by families’ juxtaposition. Subsequent comparison of relative abundances revealed five phyla common between tissues and diets. In the gut, HFD triggered an expansion of Firmicutes and a parallel reduction of Bacteroidetes. Similar changes have been detected not only in murine (Ley, Bäckhed et al. 2005) but in human gut (Ley, Turnbaugh et al. 2006), as well.

In accordance to the above described findings, relative abundance-based samples clustering proposed that during HFD liver microbiome dissociates from that of the gut and tends to equalize with that of the lungs. Four families, Staphylococcaceae, Streptococcaceae, Peptoniphilaceae and Pasteurellaceae were favored by obesity in all three tissues. Species assessment identified Staphylococcus, Streptococcus and Finegoldia as those consistently affected by HFD in all three tissues. Interestingly, we have detected several superantigen producer species, with four of them being affected by HFD in at least a single anatomic location.

Analyzing the above findings it is important to note that to the best of our knowledge, our study is the first ever to detect microbes in a healthy murine liver. These microorganisms potentially originated from the neighbor gut and reached the liver via a “leaky” portal vein (Broadley, Plaumann et al. 2016, Zeng, Surewaard et al. 2016). In agreement with our findings, microbes have recently been detected in healthy human liver (Suppli, Bagger et al. 2021). Furthermore, identification of increased abundance of Staphylococcus in all tissues during HFD, conforms to obese individuals having and increased risk of colonization from this genus (Befus, Lowy et

al. 2015). Elevated Staphylococcus load has also been associated with IPF and its acute exacerbations (Han, Zhou et al. 2014, D'Alessandro-Gabazza, Kobayashi et al. 2020). Last, the observation of super-antigen producers been affected by HFD suggests that HFD can indirectly disturb pulmonary homeostasis, similar to staphylococcal super-antigens causing interstitial pneumonia in murine models (Shinbori, Matsuki et al. 1996).

In a nutshell, we have here presented preliminary data proposing a connection between diet-induced obesity and lung fibrosis via microbiome changes. Studies of bigger sample sizes and extensive mechanistic assays are necessary for an in depth and proper statistical description of the underlying biology.

4.4 Fibrosis-related features

4.4.1 LCN2 and fibrosis

LCN2 encodes for a protein that has been functionally related to iron homeostasis (Ali, Kim et al. 2020) and microbiome regulation (Molyneaux, Cox et al. 2014), but more importantly has been detected in multiple pulmonary pathologies (Eagan, Damås et al. 2010, Kangelaris, Prakash et al. 2015, Treekitkarnmongkol, Hassane et al. 2020, Huang, Li et al. 2022). Despite the above indications a more thorough research on *LCN2* is needed for an in depth description of its actions during disease and steady state.

During the current thesis, our laboratory has investigated *LCN2* participation in the establishment and progression of IPF. To begin with, we have observed via Fibromine increased levels of *LCN2* mRNA in IPF and non-fibrotic lung tissue samples (Fanidis, Moulos et al. 2021). Increased *LCN2* expression was also negatively related to lung function as expected from a previous research deliverable (Ikezoe, Handa et al. 2014). At a single cell resolution, mRNA was primarily detected in epithelial cell types and aberrant basaloid cells, a recently described IPF-specific cellular population with spatial proximity to lung myofibroblasts (Adams, Schupp et al. 2020). Between phenotypes DEA revealed up-regulation of the gene in cell types such as AT1 and AT2, while a proper *LCN2* assessment was not possible in neutrophils, mainly due to their inadequate cells number. The aforementioned results were replicated by *LCN2* protein measurement in the BAL fluid of IPF patients.

Subsequently, we examined *Lcn2*. As with the human data, we have found an over-expression of *Lcn2* in multiple Fibromine-hosted fibrotic datasets. Re-analysis of scRNA-seq data and mining of public databases located *Lcn2* active transcription in epithelial cells and neutrophils. Interestingly, *Lcn2* mRNA quantities climax during the inflammatory stage of the animal model, connecting the gene to pre-fibrotic inflammation. Measurement of mRNA in the lung and protein in BAL fluid of BLM-treated animals revealed an up-regulation in *Lcn2* expression during both inflammatory and fibrotic phases of the model. *Lcn2* serum protein levels increased during the neutrophil-rich inflammatory phase only. Increased protein abundance was also detected in the lung during day 14 post BLM treatment. LPS-caused ALI reflected the above findings in all lung, BAL and sera, thus supporting an inflammation-related over-expression of *LCN2/Lcn2*.

Subsequently, we created *Lcn2*^{-/-} mice which were then administered BLM. Although the genetic modification did not have any significant impact on multiple BLM-triggered histopathological and physiological changes, maintenance of good lung function in treated *Lcn2*^{-/-} animals validated *Lcn2* implication in lung homeostasis. Furthermore, no major effects were observed from knocking out *Lcn2* in BLM-treated obese mice, irrespective of the previously

described triangular connection between *Lcn2*, obesity and fibrosis (Jaberi, Cohen et al. 2021, Guo, Sunil et al. 2022). Last, in contrast to the ALI model findings, LPS-induced inflammation was not impacted by *Lcn2* deficiency.

Failure of *Lcn2* KO to significantly impact BLM-treated and LPS-challenged mice can be attributed to several factors. Initially, one can hypothesize that *Lcn2* does not have any major pathogenic role, yet without neglecting *Lcn2* importance as reflected by the lung function measurements performed. On the other hand though, it may be true that *Lcn2* cannot be properly studied in the aforementioned models as previously witnessed for other genomic features (Mouratis and Aidinis 2011). It should also be noted that *Lcn2* KO may disrupt homeostasis in ways not examined during this thesis, such as via microbiome regulation (Flo, Smith et al. 2004). As far as the inflammatory role of *Lcn2* is concerned, published studies have produced contradictory reports (Jang, Lee et al. 2013, Warszawska, Gawish et al. 2013, Kang, Ren et al. 2018, Guardado, Ojeda-Juárez et al. 2021, An, Yoo et al. 2023) possibly reflecting tissue-specific *Lcn2* functions and/or differences between used models. Nevertheless, potential inability of animal models to clarify the role of *Lcn2* should not discourage future human-centered experiments.

To conclude with, we here suggest that *LCN2/Lcn2* is important for earlier inflammatory stages of lung fibrosis in both humans and mice, potentially acting via neutrophils. Future studies should address limitations of the *Lcn2*^{-/-} model in favor of a mechanistic explanation of its function.

4.4.2 TKS5 and fibroblast invasiveness

Until today, many of the IPF pathogenic and progression mechanisms are still unknown. In search of a probable disease cause, scientific community frequently shapes new hypotheses based on the common grounds between IPF and other human pathologies including cancer (Vancheri 2013).

One of the commonalities between cancer and PF is the invasive character of tissue fibroblasts (Karampitsakos, Tzilas et al. 2017), with IPF lung fibroblasts having an increased invasiveness compared to other pulmonary pathologies (Karvonen, Lehtonen et al. 2012). For the invasion to take place, fibroblasts exploit podosomes (or invadopodia in a tumor context), specialized cell structures that proteolyze ECM enabling cellular migration (Murphy and Courtneidge 2011, Paterson and Courtneidge 2018). *SH3PXD2A* (TKS5) codes for a necessary podosome component (Hoshino, Branch et al. 2013) and has previously been implicated in tumor cell metastasis including lung adenocarcinoma (Seals, Azucena et al. 2005, Li, Chen et al. 2013).

As part of this thesis, we have examined TKS5 expression in IPF using both human and murine data. To begin with, we have *in silico* observed that transcription of *TKS5* is up-regulated in multiple IPF lung samples compared to their control counterparts, while its mRNA levels correlated with those of *COL1A1*. These data were validated using both PCR and immunostaining applied on an independent sampling set. scRNA-seq data re-analysis detected *TKS5* expression in the epithelium and cells of the fibroblastic lineage. Interestingly, TKS5+ fibroblasts were positive for *COL1A1* and preferentially positive for *CD44* and *PD-L1/CD274*. CD44 is required for proper invadopodia functioning (Zhao, Xu et al. 2016) and PD-L1 has been suggested as an invasion marker (Jiang, Liu et al. 2022). Thus, TKS5+ fibroblasts should have an increased invasive potential supported by podosomes formation. In addition to human data, bulk sequencing of mouse lung tissue revealed an over-expression of *Tks5* in BLM-treated

animals. *Tks5* levels were once again correlated to those of *Colla1*. Supplementary to human tissue, immunostaining positioned *Tks5* in fibrosis-affected areas and the epithelium. Comparison of α SMA and *Colla1* expressing fibroblasts demonstrated a much more frequent co-localization of *Tks5* with the latter instead of the latter.

Subsequently, we provided proof that TGF β induces the formation of podosomes via TKS5 and MMP9 expression amplification in NHLFs and other cell lines. These results are in accordance to podosomes regulation by other growth factors, such as PDGF (Quintavalle, Elia et al. 2010, Charbonneau, Lavoie et al. 2016). On the other hand, cultured IPF-originating lung fibroblasts created podosomes without any external signal. Similarly, enhanced invadosome formation by IPF lung fibroblasts was in parallel suggested by an independent research effort (Lebel, Cliche et al. 2023). These structures lasted for a significant period of time and were highly functional. Podosomes were also developed by TGF β -treated NMLFs and untreated fibroblasts collected from the lungs of BLM-treated mice. *Tks5*, *Colla1* and *Mmp9* expression was boosted as in the case of human samples. Detection of increased MMP9/*Mmp9* levels is of specific importance as it has, along with CD44, been connected to wound healing (Michopoulou, Montmasson et al. 2020). Furthermore, it has been detected in TGF β -treated IPF lung fibroblasts of increased invasiveness (Ramírez, Hagood et al. 2011). Given the conflicting results of MMP9 targeting (Espindola, Habel et al. 2020, Bormann, Maus et al. 2022) more dedicated studies are required to illuminate its function in fibrosis.

To assess the functional importance of *Tks5* for the fibrotic phenotype we have created a haploinsufficient murine model. Genetically modified animals were sacrificed at the peak of fibrosis. Despite BLM-treatment collagen levels were found reduced in the partial absence of *Tks5*, an observation supplemented by less collagen deposition and fibrotic areas, reflected also in *ex vivo* PCLS cultures. From a physiological perspective, lung function of genetically modified animals was not that much affected compared to that of their WT counterparts. To examine the potentially anti-fibrotic actions of *Tks5*, we cultured primary lung fibroblasts from WT and *Tks5*^{+/-} animals with TGF β . The latter cells were less proliferative, with a smaller number of podosomes. Moreover, genetically modified cells had a decreased migratory potential as witnessed from transwell invasion assay. Experiments repetition using BLM instead of TGF β led to the same findings. These observations supplement previous ones supporting the invasive profile of fibrotic lung fibroblasts (White, Thannickal et al. 2003, Li, Jiang et al. 2011, Karvonen, Lehtonen et al. 2012).

From a molecular point-of-view, Quant-Seq analysis of TGF β -treated WT and *Tks5*^{+/-} lung fibroblasts revealed several thousand DEGs including multiple fibrosis-associated and invasion-related ones. Interestingly, PA pinpointed the suppression of various ECM-related processes in haploinsufficient cells indicating a functional relationship between *Tks5* and the extracellular environment. In parallel, staining of cultured BLM-treated fibroblasts showed reduced quantities of *Colla1* in *Tks5* semi-deficient cells compared to control ones. Moreover, growth of NMLFs on α ECM isolated from BLM-treated animals induced *Tks5* expression, podosomes development and *Colla1* transcription, indicating an active crosstalk between fibroblasts and the extracellular environment via their podosomes (Iizuka, Leon et al. 2020). Without neglecting the molecular compounds of post BLM ECM, podosomes formation could also have been triggered by the increased matrix stiffness, as already known for invadopodia (Alexander, Branch et al. 2008). The above hypothesis becomes even more probable if we take into consideration the inherent stiffness of lung ECM and its plasticity upon homeostasis

disruption, ageing and especially IPF (White 2015, Pardo and Selman 2016, Burgstaller, Oehrle et al. 2017).

In an attempt to repurpose drugs for podosome targeting, we have applied CMap analysis on our top Quant-Seq DEGs. Among the signatures that mimicked Tks5 knockout there was that of an Src inhibitor. Src is necessary for podosomes formation (Li, Zhao et al. 2020) and its inhibition has the strongest connectivity score to the IPF transcriptomic signature compared to nintedanib and pirfenidone (Ahangari, Becker et al. 2022). Thus, we have inhibited its function on TGF β -treated NHLFs and compared its effects with those of nintedanib. Both treatments decreased expression of *TKS5* and *COL1A1*, reduced the number of podosomes and attenuated cells' migratory potential. Incubation of mouse PCLS with A-419259 highlighted its ability to decrease fibrotic load. Last, mice that have already been treated with BLM were administered the inhibitor. Interestingly, A-419259 delivery was well tolerated and resulted in less intense pro-fibrotic hallmarks, such as collagen production and deposition. BLM-induced changes were also attenuated.

It must be noted that apart from fibroblasts, we have here shown *TKS5* transcription by epithelial and basal cells when examined at the single cell level. Podosomes *per se* have not been described in epithelial cells, yet their adhesive abilities to the basement membrane have been related to podosome-like structures (Spinardi, Rietdorf et al. 2004) as well as wound healing (Michopoulou, Montmasson et al. 2020). Src-inhibition, that was here observed to affect podosomes, has been witnessed to inhibit basal-cell mediated PF development (Jaeger, Schupp et al. 2022). Further targeted studies should be performed for demystifying *TKS5* role in these two cell types.

To conclude with, increased *TKS5* expression during fibrosis acts in a pro-fibrotic fashion by enhancing podosome formation and thus ECM invasion by lung fibroblasts. Drug targeting of the process using compounds, such as Src-inhibitors, is a promising therapeutic treatment that should be assessed by future studies.

4.4.3 MAP3K8 and fibrosis

MAP3K8 is a kinase of the MAPK pathway and is implicated in cancer (Johnson, Stuhlmiller et al. 2014) and inflammatory diseases, such as inflammatory bowel disease (Roulis, Nikolaou et al. 2014). In the pulmonary context it is protective against inflammation (Watford, Wang et al. 2009) and cancer development (Gkirtzimanaki, Gkouskou et al. 2013). In this thesis, we attempted to dissect its role in lung fibrosis, given its many commonalities with both the aforementioned pathological conditions (Vancheri 2013, Desai, Winkler et al. 2018, Savin, Zenkova et al. 2022).

Initially, measurements of *Map3k8* in BLM-treated murine lungs revealed a reduction in gene's expression. We observed the same under-expression pattern in *in silico* re-analyzed IPF versus control transcriptomic datasets. Re-analysis of scRNA-seq data pinpointed macrophages and monocytes as the main *MAP3K8* producers of the human lung, while mRNA was produced from HCs as well. Per cell type - between phenotypes DEA established a significant down-regulation of the gene in SPP1⁺ macrophage subcluster 1 and aSMA⁺ fibroblasts of IPF origin. In extension to the DEA results, IPF lung fibroblasts have been suggested to express smaller amounts of the MAP3K8-downstream molecules COX-2 and PGE2 (Wilborn, Crofford et al. 1995, Coward, Watts et al. 2009). As far as macrophages are concerned, SPP1^{hi} cells have been

located in lung fibrotic niches (Morse, Tabib et al. 2019) and were also suggested to act in a pro-fibrotic manner (Adams, Schupp et al. 2020)

For a functional examination, we created *Map3k8* knockout mice and treated them with BLM. Interestingly, multiple physiological and molecular measurements demonstrated that genetically modified animals suffered from a more severe fibrotic phenotype, suggesting that *Map3k8* exert an anti-fibrotic role in the lungs. Subsequently, we used bone marrow transplantation to and from *Map3k8*^{-/-} and WT irradiated mice in order to dissect HCs and nHCs contribution to *Map3k8*-mediated protective processes. Mice examination 14 days post-BLM administration unveiled a more severe fibrosis in animals hosting either hematopoietic or non-hematopoietic *Map3k8*-deficient cells. Our data also supported that HCs deficiency has a broader and more significant impact compared to that of nHCs. To extend our findings we specifically examined BAL fluid immune cells via FACS and discovered that *Map3k8* deletion from either HCs or nHCs affected macrophages accumulation. On the other hand, neutrophils infiltration was not affected while lymphocyte numbers were increased post to HC but not nHC genetic modification. Motivated by the potential importance of *Map3k8* for macrophages, we specifically deleted it from both macrophages and granulocytes. KO animals treated with BLM developed more severe fibrosis compared to their littermate WT controls. Interestingly, genetically modified macrophages had similar profile to the deficient HCs implicating both cell types into PF.

As MAP3K8 anti-fibrotic role may arise from its known regulation of PGE2 (Bozyk and Moore 2011, Roulis, Nikolaou et al. 2014, Roulis, Kaklamanos et al. 2020), we next examined the PGE2 pathway on the 14th day post BLM administration in WT and *Map3k8*^{-/-} animals. *Cox-2* expression was down-regulated in KO mice, while PGE2 and AA abundance was reduced in the BALF, thus supplementing previous reports in other pathological contexts (Eliopoulos, Dumitru et al. 2002, Berthou, Ceppo et al. 2015). On the other hand, animal treatment with a PGE2 analogue partially rescued the *Map3k8*^{-/-} phenotype and reduced the fibrosis load of WT controls. Last, treatment with PGE2 *per se* constricted fibrosis in BLM-treated animals as expected (Failla, Genovese et al. 2009, Dackor, Cheng et al. 2011), further supporting the importance of the MAP3K8-PGE2 pathway for PF. Implication of PGE2 in the apoptosis and differentiation of lung fibroblast (Maher, Evans et al. 2010, Bozyk and Moore 2011) could at least partially explain the phenotype of animals with *Map3k8* deficient nHCs.

Conclusively, we have shown that lung-expressed *Map3k8* acts in an anti-fibrotic manner via the *Map3k8/Cox-2/PGE2* axis. During fibrosis its expression is reduced in specific cell types, such as macrophages and fibroblasts. Opposite findings (Perugorria, Murphy et al. 2013) may reflect organ-specific processes, as suggested elsewhere (Chen and Dai 2023). Future studies are necessary for distinction of fibrotic and inflammatory *Map3k8/Cox-2/PGE2* effects, as well as for the examination of other *Map3k8* pathways of action.

4.5 Fibrosis and COVID-19

SARS-CoV-2-caused COVID-19 is a disease initially detected in late 2019 and since then have affected millions of individuals worldwide (Hu, Guo et al. 2021). It can be accompanied by a plethora of symptoms defined by each patient's underlying clinical picture, but is usually characterized by hypoxaemia and even ARDS. COVID-19 is closely related to PF with hallmarks of the latter witnessed in severe cases of the former (Osuchowski, Winkler et al.

2021) and ILD patients having a less favorable COVID-19 progression if infected (Drake, Docherty et al. 2020, Esposito, Menon et al. 2020, Beltramo, Cottenet et al. 2021).

During this thesis we have explored the relationship of COVID-19 and IPF at a molecular level via examination of the pro-fibrotic molecule *ENPP2*. Initially, we have observed increased *ENPP2* levels (either mRNA or protein) in the nasopharynx and blood serum of COVID-19 patients. In detail, ATX levels were potentially connected to disease severity as Dex-untreated ICU patients had higher protein quantities related to individuals treated in the WARD. There was also not any association between significant ATX changes and patients' clinical characteristics. Subsequently, extending findings in ARDS suffering individuals (Gao, Li et al. 2021), we have also demonstrated that IL-6 serum levels correlated with those of ATX in ICU patients not treated with Dex, suggesting an *ENPP2* implication in severe COVID-19 cases. ATX levels co-fluctuated with epithelium-related proteins, as well. As expected from previous experiments (Meng, Tang et al. 2019), measurements from two independent cohorts showed that ATX was less abundant in Dex-treated ICU residents with mechanical ventilation compared to non-treated ventilated counterparts. Remarkably, ATX levels were the highest in samples from deceased patients that had not been treated with Dex. No difference was observed in non-ventilated individuals.

Afterwards, we examined *ENPP2* mRNA levels at the single cell level. Re-analysis of eight scRNA-seq datasets detected active *ENPP2* transcription in immune cells of the nasopharynx and pDCs of the peripheral circulation. pDCs were the main *ENPP2* producers among BAL cells, too. On the other hand, *ENPP2* was expressed by mesothelial and arterial cells, as well as by monocytes located in COVID-19 patients' lungs. Interestingly, lung arterial cells from IPF individuals were the main source of ATX mRNA along with fibroblast subtypes and DCs. Despite the small number of the latter, we established a significant difference in *ENPP2* mRNA levels between IPF and healthy samples-originating peripheral pDCs. The same difference was observed between lung DCs sampled from infected and non-infected lungs. As far as both lung and peripheral DCs are concerned, *ENPP2* mRNA was not detected in a big number of cells suggesting the existence of an either *ENPP2*+ subset or a dropout effect in place.

In validation of the scRNA-seq data presented above, *ENPP2* transcription was found up-regulated in pDCs of various inflammatory diseases compared to cells arising from steady state tissues according to bulk RNA-seq data. During homeostasis, pDCs were the primary supplier of the immune compartment with *ENPP2* and *LPAR2*. Last, single cell level DEA highlighted *TMEM176B* and *CD1a* as the only genomic features of DCs that co-express with *ENPP2* during COVID-19. Strikingly, it has been previously suggested that both features mark early stages in DC development (Tazi, Bouchonnet et al. 1993, Condamine, Le Texier et al. 2010, Bourgeois, Subramaniam et al. 2015, Picotto, Morse et al. 2019), thus proposing a new role for *ENPP2* as regulator of DC homeostasis.

On the basis of both the above findings, we are of the opinion that *ENPP2* over-expression in COVID-19 pDCs should be evaluated along with the observation of a diminished blood pDCs population during COVID-19 (Zhou, To et al. 2020). Developmentally, pDCs are controlled by features such as *TMEM176B*, *CD1a* – see previous paragraph – and *TCF4* (Cisse, Caton et al. 2008) that is evidently modulated by LPA (Guo, He et al. 2015). Thus, we suggest that increased autocrine ATX signaling during SARS-CoV-2 infection could prevent maturation of

pDCs leading to the aforementioned depletion. In support of our hypothesis, LPA anti-inflammatory functions have previously been described (Emo, Meednu et al. 2012).

Another possible functional relationship between ATX and COVID-19 comes through vasculature homeostasis. *ENPP2* was here found to be expressed by cells of the pulmonary circulation and was elsewhere connected with coronaviruses-caused lung hemorrhages (Gralinski, Ferris et al. 2015, Gralinski, Menachery et al. 2017). It has also been proposed that ATX-LPA axis stimulates E-sel expression in human endothelial cells (Rizza, Leitinger et al. 1999, Shlyonsky, Naeije et al. 2014, Takeda, Matoba et al. 2019). Given the fact that ATX has positively correlated with the increased levels of soluble E-selectin in non-survivors of the Evangelismos cohort (Vassiliou, Keskinidou et al. 2021), it is probable that ATX participates in the endothelial impairment seen during COVID-19 (Flaumenhaft, Enjyoji et al. 2022).

In conclusion, we have demonstrated that both ATX mRNA and protein levels are increased during SARS-CoV-2 infection in the periphery and the lungs. We also propose that ATX can affect DCs maturation and thus the inflammatory response of the patients, as well as their vasculature homeostasis. Future publications examining larger numbers of DCs and applying further wet laboratory techniques are necessary for in depth examination of these novel hypotheses.

4.6 LPA in the kidneys

LPA and another 175 stimulants have been used to treat HKC-8 cells in order to measure their effect on protein secretion and phosphorylation. According to a custom multiplex ELISA, an active response was recorded in %61 of the stimulants, with HGF inducing the greatest number of changes. Stimulants clustering revealed three major groups, each one with distinct impact on kidney tubular epithelial cells. Specifically, LPA 18:1 was a cluster II member along with several pro-inflammatory factors, such as PMA, IL1A-B, TNFA and IL17A, suggesting an active role in CKD. Cell culturing with three different LPA species up-regulated phosphorylation of I κ Ba, MEK1, CREB1, ERK1 (marginally) and JUN, the most activated signal transduction molecule of our experiment. The same treatment induced the secretion of CCL2, CCL3, CXCL10, ICAM1, IL-6, IL-8 and CCL3. Respective genes' transcription up-regulation was also witnessed. In accordance to previous publications (Pradere, Klein et al. 2007, Geng, Lan et al. 2012, Sakai, Chun et al. 2017), LPA effects on kidney tubular epithelial cells seem to be primarily manifested via LPAR2 and secondarily via LPAR6.

In more detail, CCL2-3 are known for their chemo-attraction of various inflammatory cells (Lv, Booz et al. 2018) and CCL2 has also been implicated in various renal-related pathologies and animal models (Wada, Yokoyama et al. 1996, Lloyd, Minto et al. 1997, Vielhauer, Anders et al. 2001, de Zeeuw, Bekker et al. 2015). On the contrary, although a known chemoattractant (Vazirinejad, Ahmadi et al. 2014) CXCL10 role in CKD requires further research. ICAM1 is detected in kidney tubular epithelial cells specifically during pathological conditions such as IgA nephropathy (Arrizabalaga, Solé et al. 2003) and primary glomerulosclerosis (Vazirinejad, Ahmadi et al. 2014). Re-analysis of a public microarray dataset validated the pathogenic role of ICAM1 with its transcription being up-regulated during IgA nephropathy, CKD and focal segmental glomerulosclerosis compared to steady state. Although not targeted by LPA, TNFSF12, IL-1 β and CCL5 cluster II induced molecules were also among the deregulated features. Similarly to ICAM1, IL-6 levels are higher in CKD than in control samples (Oh, Kim et al. 2013), while IL-6 has also been implicated in fibrosis via collagen I induction

(Ranganathan, Jayakumar et al. 2013). IL-8 has been previously reported downstream of LPA in the lungs (Cummings, Zhao et al. 2004, Barezzi, Roman et al. 2006, Saatian, Zhao et al. 2006) and its levels have been found altered in the kidneys of type-2 diabetes (Loretelli, Rocchio et al. 2021) and CKD patients (Tunçay, Doğan et al. 2021). JUN and CREB1 are two TFs responsible for regulating the expression of various inflammatory-related genes including the LPA-induced *IL-6* (Wen, Sakamoto et al. 2010, Grynberg, Ma et al. 2017), while MEK1 and ERK – members of the same pathway – are known targets in the lung context (Saatian, Zhao et al. 2006, Nathan, Zhang et al. 2021). Last, phosphorylation of I κ B α upon LPA treatment induces the expression of NF κ B, a well-known player in various immune-related pathologies (Haij, Woltman et al. 2002, Hayden and Ghosh 2012). In total and with the exception of CCL2 and ERK, our experiment showed for the first time the LPA induction of the above molecules' secretion/phosphorylation from/in kidney tubular epithelial cells.

Apart from their importance under pathological conditions, LPA-induced TFs can also regulate the expression of many LPA-affected molecules that we measured. Use of an NF κ B-specific inhibitor blocked LPA-mediated activation of CCL2, CCL3, CXCL-8, and ICAM1 extending previous findings (Melotti, Nicolis et al. 2001, Viedt and Orth 2002, Cummings, Zhao et al. 2004, Brasier 2010, Deng, Xu et al. 2013). On the other hand and although known to be a target of NF κ B pathway (Georganas, Liu et al. 2000, Liu, Zhang et al. 2017) IL-6 was not affected, probably due to cell type specific reasons. Our results also validate the capability of human primary proximal TECs to produce IL-6, IL-8, and CCL2. JNK was witnessed to regulate ICAM1 and CCL3 abundance in accordance to previous publications (Holzberg, Knight et al. 2003, Sindhu, Akhter et al. 2020), while no effect was observed for CCL2 and CXCL8. Increased CREB1 phosphorylation post LPA treatment could explain LPA impact on CXCL8, CCL2 and CCL3 according to the existing literature (Mayer, Simard et al. 2013, Zhao, Hui et al. 2014, Koga, Tsurumaki et al. 2019, Tomalka, Pelletier et al. 2021). Last, CCL2, CCL3, CXCL10, ICAM1, IL-6, CXCL8 and CCL3 expression could be regulated by MAPK signaling. Our ELISA measurements supported such regulation for *CCL2*, *CCL3*, *CXCL8*, and *ICAM1* in line with previous findings (Hwang, Jeong et al. 2004, Lee, Zhang et al. 2006, Zhu, Wei et al. 2016, Namba, Nakano et al. 2017), while IL-6 abundance remained unaffected adding to an existing list of conflicting results (Klemm, Bruchhagen et al. 2017, Kitanaka, Nakano et al. 2019, Hu, Huang et al. 2020, Plastira, Bernhart et al. 2020). We did not test CXCL10 expression dependence on MAPK pathway.

As far as the other cluster II members are concerned, most of them are positively related with CKD progression. Indicatively, TNFSF12/TWEAK levels significantly diverge during renal injury and related pathologies compared to steady state (Ortiz, Sanz et al. 2009, Ucerro, Benito-Martin et al. 2013, Sanz, Izquierdo et al. 2014, Bernardi, Voltan et al. 2019) and TNFSF12/TWEAK inhibition reduces inflammation and partially restores kidney function (Sanz, Justo et al. 2008, Poveda, Tabara et al. 2013, Bernardi, Voltan et al. 2019). TNF α is directly related to renal function as its very low homeostatic levels are necessary for kidneys' excretory functions to take place (Mehaffey and Majid 2017). Up-regulation of its expression has been witnessed during various kidney pathologies (Vielhauer and Mayadas 2007, Nikitopoulou, Oikonomou et al. 2012). IL-1 β induces inflammation and causes IgA nephropathy (Stangou, Papagianni et al. 2013), while it is secreted by kidney cells during tissue injury alongside IL-1 α (Anders 2016). On the contrary, findings are rather controversial for IL-17a (Ramani, Tan et al. 2018, Sun, Wang et al. 2018, Orejudo, Rodrigues-Diez et al. 2019).

The above observations suggest that LPA functions in a pro-pathogenic manner in the kidneys since it clusters with CKD-promoting molecules. Indeed, available literature supports our hypothesis (Pradere, Klein et al. 2007, Swaney, Chapman et al. 2011, Mirzoyan, Baiotto et al. 2016, Sakai, Chun et al. 2017).

In conclusion, our study has provided detailed information regarding human renal proximal TECs' response to a great variety of stimulants. In specific, LPA was shown to exert a pro-inflammatory function leading in the regulation of major signaling pathways and the secretion of multiple molecules potentially affecting kidney homeostatic maintenance.

References

1. Abram, C. L., D. F. Seals, I. Pass, D. Salinsky, L. Maurer, T. M. Roth and S. A. Courtneidge (2003). "The Adaptor Protein Fish Associates with Members of the ADAMs Family and Localizes to Podosomes of Src-transformed Cells *." *Journal of Biological Chemistry* 278(19): 16844-16851.
2. Adams, T. S., J. C. Schupp, S. Poli, E. A. Ayaub, N. Neumark, F. Ahangari, S. G. Chu, B. A. Raby, G. DeJuliis, M. Januszyk, Q. Duan, H. A. Arnett, A. Siddiqui, G. R. Washko, R. Homer, X. Yan, I. O. Rosas and N. Kaminski (2020). "Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis." *Science Advances* 6(28): eaba1983.
3. Agassandian, M., J. R. Tedrow, J. Sembrat, D. J. Kass, Y. Zhang, E. A. Goncharova, N. Kaminski, R. K. Mallampalli and L. J. Vuga (2015). "VCAM-1 is a TGF- β 1 inducible gene upregulated in idiopathic pulmonary fibrosis." *Cellular Signalling* 27(12): 2467-2473.
4. Ahangari, F., C. Becker, D. G. Foster, M. Chioccioli, M. Nelson, K. Beke, X. Wang, A. Justet, T. Adams, B. Readhead, C. Meador, K. Correll, L. N. Lili, H. M. Roybal, K.-A. Rose, S. Ding, T. Barnthaler, N. Briones, G. DeJuliis, J. C. Schupp, Q. Li, N. Omote, Y. Aschner, L. Sharma, K. W. Kopf, B. Magnusson, R. Hicks, A. Backmark, C. S. Dela Cruz, I. Rosas, L. P. Cousens, J. T. Dudley, N. Kaminski and G. P. Downey (2022). "Saracatinib, a Selective Src Kinase Inhibitor, Blocks Fibrotic Responses in Preclinical Models of Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 206(12): 1463-1479.
5. Ahmadvand, N., G. Carraro, M. R. Jones, I. Shalashova, A. Noori, J. Wilhelm, N. Baal, F. Khosravi, C. Chen, J.-S. Zhang, C. Ruppert, A. Guenther, R. M. Wasnick and S. Bellusci (2022) "Cell-Surface Programmed Death Ligand-1 Expression Identifies a Sub-Population of Distal Epithelial Cells Enriched in Idiopathic Pulmonary Fibrosis." *Cells* 11 DOI: 10.3390/cells11101593.
6. Ahmadvand, N., A. Lingampally, F. Khosravi, A. I. Vazquez-Armendariz, S. Rivetti, M. R. Jones, J. Wilhelm, S. Herold, G. Barreto, J. Koepke, C. Samakovlis, G. Carraro, J.-S. Zhang, D. Al Alam and S. Bellusci (2022). "Fgfr2b signaling is essential for the maintenance of the alveolar epithelial type 2 lineage during lung homeostasis in mice." *Cellular and Molecular Life Sciences* 79(6): 302.
7. Alexander, N. R., K. M. Branch, A. Parekh, E. S. Clark, I. C. Iwueke, S. A. Guelcher and A. M. Weaver (2008). "Extracellular Matrix Rigidity Promotes Invadopodia Activity." *Current Biology* 18(17): 1295-1299.
8. Ali, M. K., R. Y. Kim, A. C. Brown, C. Donovan, K. S. Vanka, J. R. Mayall, G. Liu, A. L. Pillar, B. Jones-Freeman, D. Xenaki, T. Borghuis, R. Karim, J. W. Pinkerton, R. Aryal, M. Heidari, K. L. Martin, J. K. Burgess, B. G. Oliver, D. Trinder, D. M. Johnstone, E. A. Milward, P. M. Hansbro and J. C. Horvat (2020). "Critical role for iron accumulation in the pathogenesis of fibrotic lung disease." *The Journal of Pathology* 251(1): 49-62.
9. Allen, R. J., J. Porte, R. Braybrooke, C. Flores, T. E. Fingerlin, J. M. Oldham, B. Guillen-Guio, S.-F. Ma, T. Okamoto, A. E. John, M. e. Obeidat, I. V. Yang, A. Henry, R. B. Hubbard, V. Navaratnam, G. Saini, N. Thompson, H. L. Booth, S. P. Hart, M. R. Hill, N. Hirani, T. M. Maher, R. J. McAnulty, A. B. Millar, P. L. Molyneaux, H. Parfrey, D. M. Rassl, M. K. B. Whyte, W. A. Fahy, R. P. Marshall, E. Oballa, Y. Bossé, D. C. Nickle, D. D. Sin, W. Timens, N. Shrine, I. Sayers, I. P. Hall, I. Noth, D. A. Schwartz, M. D. Tobin, L. V. Wain and R. G. Jenkins (2017). "Genetic variants associated with

- susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study." *The Lancet Respiratory Medicine* 5(11): 869-880.
10. An, H. S., J.-W. Yoo, J. H. Jeong, M. Heo, S. H. Hwang, H. M. Jang, E. A. Jeong, J. Lee, H. J. Shin, K. E. Kim, M. C. Shin and G. S. Roh (2023). "Lipocalin-2 promotes acute lung inflammation and oxidative stress by enhancing macrophage iron accumulation." *International Journal of Biological Sciences* 19(4): 1163-1177.
 11. Anders, H.-J. (2016). "Of Inflammasomes and Alarmins: IL-1 β and IL-1 α in Kidney Disease." *Journal of the American Society of Nephrology* 27(9).
 12. Anders, S. and W. Huber (2010). "Differential expression analysis for sequence count data." *Genome biology* 11(10): R106-R106.
 13. Anderson, A. L., Y. Zheng, D. Song, D. LaRosa, N. Van Rooijen, G. Kierstein, S. Kierstein, A. Haczku and A. I. Levinson (2012). "The B-cell superantigen *Fingoldia magna* protein L causes pulmonary inflammation by a mechanism dependent on MyD88 but not B cells or immunoglobulins." *Inflammation Research* 61(2): 161-169.
 14. Antoniou, K. M., E. Tsitoura, E. Vasarmidi, E. K. Symvoulakis, V. Aidinis, V. Tzilas, A. Tzouvelekis and D. Bouros (2021). "Precision medicine in idiopathic pulmonary fibrosis therapy: From translational research to patient-centered care." *Current Opinion in Pharmacology* 57: 71-80.
 15. Aran, D., A. P. Looney, L. Liu, E. Wu, V. Fong, A. Hsu, S. Chak, R. P. Naikawadi, P. J. Wolters, A. R. Abate, A. J. Butte and M. Bhattacharya (2019). "Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage." *Nature Immunology* 20(2): 163-172.
 16. Arcos-Montoya, D., T. Wegman-Ostrosky, S. Mejía-Pérez, M. De la Fuente-Granada, I. Camacho-Arroyo, A. García-Carrancá, M. Velasco-Velázquez, J. Manjarrez-Marmolejo and A. González-Arenas (2021). "Progesterone Receptor Together with PKC α Expression as Prognostic Factors for Astrocytomas Malignancy." *OncoTargets and Therapy* 14: 3757-3768.
 17. Arrizabalaga, P., M. Solé, R. Abellana, X. de las Cuevas, J. Soler, J. Pascual and C. Ascaso (2003). "Tubular and Interstitial Expression of ICAM-1 as a Marker of Renal Injury in IgA Nephropathy." *American Journal of Nephrology* 23(3): 121-128.
 18. Arteel, G. E. (2020). "Liver-lung axes in alcohol-related liver disease." *Clin Mol Hepatol* 26(4): 670-676.
 19. Badgeley, M. A., S. C. Sealfon and M. D. Chikina (2015). "Hybrid Bayesian-rank integration approach improves the predictive power of genomic dataset aggregation." *Bioinformatics* 31(2): 209-215.
 20. Barbayianni, I., P. Kanellopoulou, D. Fanidis, D. Nastos, E.-D. Ntouskou, A. Galaris, V. Harokopos, P. Hatzis, E. Tsitoura, R. Homer, N. Kaminski, K. M. Antoniou, B. Crestani, A. Tzouvelekis and V. Aidinis (2023). "SRC and TKS5 mediated podosome formation in fibroblasts promotes extracellular matrix invasion and pulmonary fibrosis." *Nature Communications* 14(1): 5882.
 21. Barbayianni, I., P. Kanellopoulou, D. Fanidis, E.-D. Ntouskou, D. Nastos, A. Galaris, V. Harokopos, P. Hatzis, E. Tsitoura, R. Homer, N. Kaminski, K. M. Antoniou, B. Crestani, A. Tzouvelekis and V. Aidinis (2023). "SRC-mediated and TKS5-enabled podosome formation is an inherent property of IPF fibroblasts, promoting ECM invasion and pulmonary fibrosis." *bioRxiv*: 2023.2001.2025.522705.
 22. Barbayianni, I., I. Ninou, A. Tzouvelekis and V. Aidinis (2018). "Bleomycin Revisited: A Direct Comparison of the Intratracheal Micro-Spraying and the Oropharyngeal Aspiration Routes of Bleomycin Administration in Mice." *Frontiers in Medicine* 5: 269.
 23. Barekzi, E., J. Roman, K. Hise, S. Georas and J. W. Steinke (2006). "Lysophosphatidic acid stimulates inflammatory cascade in airway epithelial cells." *Prostaglandins, Leukotrienes and Essential Fatty Acids* 74(6): 357-363.
 24. Barnes, H., S. M. Humphries, P. M. George, D. Assayag, I. Glaspole, J. A. Mackintosh, T. J. Corte, M. Glassberg, K. A. Johannson, L. Calandriello, F. Felder, A. Wells and S. Walsh (2023). "Machine

- learning in radiology: the new frontier in interstitial lung diseases." *The Lancet Digital Health* 5(1): e41-e50.
25. Barrett, T., T. O. Suzek, D. B. Troup, S. E. Wilhite, W.-C. Ngau, P. Ledoux, D. Rudnev, A. E. Lash, W. Fujibuchi and R. Edgar (2005). "NCBI GEO: mining millions of expression profiles—database and tools." *Nucleic Acids Research* 33(suppl_1): D562-D566.
 26. Barrett, T., S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C. L. Robertson, N. Serova, S. Davis and A. Soboleva (2013). "NCBI GEO: archive for functional genomics data sets—update." *Nucleic Acids Research* 41(D1): D991-D995.
 27. Bauer, Y., J. Tedrow, S. de Bernard, M. Birker-Robaczewska, K. F. Gibson, B. J. Guardela, P. Hess, A. Klenk, K. O. Lindell, S. Poirey, B. Renault, M. Rey, E. Weber, O. Nayler and N. Kaminski (2014). "A Novel Genomic Signature with Translational Significance for Human Idiopathic Pulmonary Fibrosis." *American Journal of Respiratory Cell and Molecular Biology* 52(2): 217-231.
 28. Baumgartner, K. B., J. M. Samet, D. B. Coultas, C. A. Stidley, W. C. Hunt, T. V. Colby, J. A. Waldron and C. Collaborating (2000). "Occupational and Environmental Risk Factors for Idiopathic Pulmonary Fibrosis: A Multicenter Case-Control Study." *American Journal of Epidemiology* 152(4): 307-315.
 29. Becht, E., L. McInnes, J. Healy, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux and E. W. Newell (2019). "Dimensionality reduction for visualizing single-cell data using UMAP." *Nature Biotechnology* 37(1): 38-44.
 30. Befus, M., F. D. Lowy, B. A. Miko, D. V. Mukherjee, C. T. A. Herzig and E. L. Larson (2015). "Obesity as a Determinant of *Staphylococcus aureus* Colonization Among Inmates in Maximum-Security Prisons in New York State." *American Journal of Epidemiology* 182(6): 494-502.
 31. Beisang, D. J., K. Smith, L. Yang, A. Benyumov, A. Gilbertsen, J. Herrera, E. Lock, E. Racila, C. Forster, B. J. Sandri, C. A. Henke and P. B. Bitterman (2020). "Single-cell RNA sequencing reveals that lung mesenchymal progenitor cells in IPF exhibit pathological features early in their differentiation trajectory." *Scientific Reports* 10(1): 11162.
 32. Belkaid, Y. and Timothy W. Hand (2014). "Role of the Microbiota in Immunity and Inflammation." *Cell* 157(1): 121-141.
 33. Beltramo, G., J. Cottenet, A.-S. Mariet, M. Georges, L. Piroth, P. Tubert-Bitter, P. Bonniaud and C. Quantin (2021). "Chronic respiratory diseases are predictors of severe outcome in COVID-19 hospitalised patients: a nationwide study." *European Respiratory Journal* 58(6): 2004474.
 34. Benesch, M. G. K., X. Tang, T. Maeda, A. Ohhata, Y. Y. Zhao, B. P. C. Kok, J. Dewald, M. Hitt, J. M. Curtis, T. P. W. McMullen and D. N. Brindley (2014). "Inhibition of autotaxin delays breast tumor growth and lung metastasis in mice." *The FASEB Journal* 28(6): 2655-2666.
 35. Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1): 289-300.
 36. Bernardi, S., R. Voltan, E. Rimondi, E. Melloni, D. Milani, C. Cervellati, D. Gemmati, C. Celeghini, P. Secchiero, G. Zauli and V. Tisato (2019). "TRAIL, OPG, and TWEAK in kidney disease: biomarkers or therapeutic targets?" *Clinical Science* 133(10): 1145-1166.
 37. Berthou, F., F. Ceppo, K. Dumas, F. Massa, B. Vergoni, S. Alemany, M. Cormont and J.-F. Tanti (2015). "The Tpl2 Kinase Regulates the COX-2/Prostaglandin E2 Axis in Adipocytes in Inflammatory Conditions." *Molecular Endocrinology* 29(7): 1025-1036.
 38. Bharat, A., M. Querrey, N. S. Markov, S. Kim, C. Kurihara, R. Garza-Castillon, A. Manerikar, A. Shilatifard, R. Tomic, Y. Politanska, H. Abdala-Valencia, A. V. Yeldandi, J. W. Lomasney, A. V. Misharin and G. R. S. Budinger (2020). "Lung transplantation for patients with severe COVID-19." *Science Translational Medicine* 12(574): eabe4282.

39. Blouw, B., M. Patel, S. Iizuka, C. Abdullah, W. K. You, X. Huang, J.-L. Li, B. Diaz, W. B. Stallcup and S. A. Courtneidge (2015). "The Invadopodia Scaffold Protein Tks5 Is Required for the Growth of Human Breast Cancer Cells In Vitro and In Vivo." *PLOS ONE* 10(3): e0121003.
40. Blouw, B., D. F. Seals, I. Pass, B. Diaz and S. A. Courtneidge (2008). "A role for the podosome/invadopodia scaffold protein Tks5 in tumor growth in vivo." *European Journal of Cell Biology* 87(8): 555-567.
41. Bormann, T., R. Maus, J. Stolper, M. Tort Tarrés, C. Brandenberger, D. Wedekind, D. Jonigk, T. Welte, J. Gauldie, M. Kolb and U. A. Maus (2022). "Role of matrix metalloprotease-2 and MMP-9 in experimental lung fibrosis in mice." *Respiratory Research* 23(1): 180.
42. Bourgeois, E. A., S. Subramaniam, T.-Y. Cheng, A. De Jong, E. Layre, D. Ly, M. Salimi, A. Legaspi, R. L. Modlin, M. Salio, V. Cerundolo, D. B. Moody and G. Ogg (2015). "Bee venom processes human skin lipids for presentation by CD1a." *Journal of Experimental Medicine* 212(2): 149-163.
43. Bowman, W. S., G. A. Echt and J. M. Oldham (2021). "Biomarkers in Progressive Fibrosing Interstitial Lung Disease: Optimizing Diagnosis, Prognosis, and Treatment Response." *Frontiers in Medicine* 8.
44. Bozyk, P. D. and B. B. Moore (2011). "Prostaglandin E2 and the Pathogenesis of Pulmonary Fibrosis." *American Journal of Respiratory Cell and Molecular Biology* 45(3): 445-452.
45. Bradley, G. and S. J. Barrett (2017). "CausalR: extracting mechanistic sense from genome scale data." *Bioinformatics* 33(22): 3670-3672.
46. Brasier, A. R. (2010). "The nuclear factor- κ B–interleukin-6 signalling pathway mediating vascular inflammation." *Cardiovascular Research* 86(2): 211-218.
47. Brazma, A., P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansoorge, C. A. Ball, H. C. Causton, T. Gaasterland, P. Glenisson, F. C. P. Holstege, I. F. Kim, V. Markowitz, J. C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo and M. Vingron (2001). "Minimum information about a microarray experiment (MIAME)—toward standards for microarray data." *Nature Genetics* 29(4): 365-371.
48. Brisbin, A. G., Y. W. Asmann, H. Song, Y.-Y. Tsai, J. A. Aakre, P. Yang, R. B. Jenkins, P. Pharoah, F. Schumacher, D. V. Conti, D. J. Duggan, M. Jenkins, J. Hopper, S. Gallinger, P. Newcomb, G. Casey, T. A. Sellers and B. L. Fridley (2011). "Meta-analysis of 8q24 for seven cancers reveals a locus between NOV and ENPP2 associated with cancer development." *BMC Medical Genetics* 12(1): 156.
49. Broadley, Steven P., A. Plaumann, R. Coletti, C. Lehmann, A. Wanisch, A. Seidlmeier, K. Esser, S. Luo, Patrick C. Rämer, S. Massberg, Dirk H. Busch, M. van Lookeren Campagne and A. Verschoor (2016). "Dual-Track Clearance of Circulating Bacteria Balances Rapid Restoration of Blood Sterility with Induction of Adaptive Immunity." *Cell Host & Microbe* 20(1): 36-48.
50. Buechler, M. B., R. N. Pradhan, A. T. Krishnamurty, C. Cox, A. K. Calviello, A. W. Wang, Y. A. Yang, L. Tam, R. Caothien, M. Roose-Girma, Z. Modrusan, J. R. Arron, R. Bourgon, S. Müller and S. J. Turley (2021). "Cross-tissue organization of the fibroblast lineage." *Nature* 593(7860): 575-579.
51. Bueno, M., Y.-C. Lai, Y. Romero, J. Brands, C. M. St. Croix, C. Kamga, C. Corey, J. D. Herazo-Maya, J. Sembrat, J. S. Lee, S. R. Duncan, M. Rojas, S. Shiva, C. T. Chu and A. L. Mora (2015). "PINK1 deficiency impairs mitochondrial homeostasis and promotes lung fibrosis." *The Journal of Clinical Investigation* 125(2): 521-538.
52. Burgstaller, G., B. Oehrle, M. Gerckens, S. E. White, B. H. Schiller and O. Eickelberg (2017). "The instructive extracellular matrix of the lung: basic composition and alterations in chronic lung disease." *European Respiratory Journal* 50(1): 1601805.
53. Burkov, A. (2019). *The hundred-page machine learning book*, Burkov, Andriy.

54. Cabello-Aguilar, S., M. Alame, F. Kon-Sun-Tack, C. Fau, M. Lacroix and J. Colinge (2020). "SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics." *Nucleic Acids Research* 48(10): e55-e55.
55. Cai, L., J. Rubin, W. Han, P. Venge and S. Xu (2010). "The Origin of Multiple Molecular Forms in Urine of HNL/NGAL." *Clinical Journal of the American Society of Nephrology* 5(12).
56. Calfee, C. S., K. L. Delucchi, P. Sinha, M. A. Matthay, J. Hackett, M. Shankar-Hari, C. McDowell, J. G. Laffey, C. M. O'Kane, D. F. McAuley, A. J. Johnston, A. Paikray, C. Yates, P. Polgarova, E. Price, A. McInerney, K. Zamoscik, G. Dempsey, C. Seasman, L. Gilfeather, N. Hemmings, S. O'Kane, P. Johnston, L. Pokorny, C. Nutt, O. O'Neill, P. Prashast, C. Smalley, R. Jacob, J. O'Rourke, S. F. Sultan, C. Schilling, G. D. Perkins, T. Melody, K. Couper, R. Daniels, F. Gao, J. Hull, T. Gould, M. Thomas, K. Sweet, D. Breen, E. Neau, W. J. Peel, C. Jardine, P. Jefferson, S. E. Wright, K. Harris, M. Thomas, S. Hierons, J. Laffey, V. McInerney, L. Camporota, K. Lei, S. Kaul, M. Chibvuri, A. Gratrix, R. Bennett, V. Martinson, L. Sleight, N. Smith, P. A. Hopkins, D. Hadfield, S. Casboul, F. Wade-Smith, J. Dawson, C. Mellis, C. Harris, G. Parsons, S. Helyar, A. R. Bodenham, S. Elliot, Z. Beardow, S. Birch, B. Marsh, T. Martin, A. Dhrampal, M. Rosbergen, S. Webb, F. Bottrill, H. Reschreiter, H. Barcraft-Barnes, J. Camsooksai, A. Johnston, A. Clarkson, C. Bentley, L. Cooper, Y. Qui, N. Mitchell, R. Carrera, A. Whitehouse, C. M. Danbury, N. Jacques, A. Brown, D. Rogerson, C. Morris, T. Walsh, M. Gillies, G. Price, K. Kefala, N. Young, D. Hope, C. McCulloch, J. Antonelli, P. Ramsay, K. Everingham, L. Boardman, H. Dawson, F. Pollock, J. Thompson, I. D. Welters, L. Poole, P. Hampshire, A. Hall, K. Williams, A. Walker, L. Youds, S. Hendry, V. Waugh, J. Patrick-Heselton, D. Shaw, I. Chaudry, J. Baldwin, S. Drage, L. Ortiz-Ruiz de Gordo, D. McAuley, L. Bannon, V. Quinn, L. McNamee, G. White, M. Cecconi, J. Mellishoff, D. Ryan, A. Nichol, B. Agarwal, P. Meale, S. James, K. Dhadwal, D. Martin, A. Walecka, S. Ward, J. Trinder, S. Hagan, J. Montgomery, C. Leonard, E. Lemon, T. Trinick, M. Buddhavarapu, G. Ward, C. Bassford, A. Davidson, K. McGuigan, A. Benchiheb, N. Hickey, A. Binning, S. Henderson, J. A. Wood, A. J. Burtenshaw, D. Kelly, T. Martin, J. Thrush, J. Wollaston, S. Graystone, G. Nicol and G. Sellors (2018). "Acute respiratory distress syndrome subphenotypes and differential response to simvastatin: secondary analysis of a randomised controlled trial." *The Lancet Respiratory Medicine* 6(9): 691-698.
57. Callahan, B. J., P. J. McMurdie and S. P. Holmes (2017). "Exact sequence variants should replace operational taxonomic units in marker-gene data analysis." *The ISME Journal* 11(12): 2639-2643.
58. Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson and S. P. Holmes (2016). "DADA2: High-resolution sample inference from Illumina amplicon data." *Nature Methods* 13(7): 581-583.
59. Carlson, M. R. J., B. Zhang, Z. Fang, P. S. Mischel, S. Horvath and S. F. Nelson (2006). "Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks." *BMC Genomics* 7(1): 40.
60. Carraro, G., A. Mulay, C. Yao, T. Mizuno, B. Konda, M. Petrov, D. Lafkas, J. R. Arron, C. M. Hogaboam, P. Chen, D. Jiang, P. W. Noble, S. H. Randell, J. L. McQualter and B. R. Stripp (2020). "Single-Cell Reconstruction of Human Basal Cell Diversity in Normal and Idiopathic Pulmonary Fibrosis Lungs." *American Journal of Respiratory and Critical Care Medicine* 202(11): 1540-1550.
61. Carvalho, B. S. and R. A. Irizarry (2010). "A framework for oligonucleotide microarray preprocessing." *Bioinformatics* 26(19): 2363-2367.
62. Castelino, F. V., G. Bain, V. A. Pace, K. E. Black, L. George, C. K. Probst, L. Goulet, R. Lafyatis and A. M. Tager (2016). "An Autotaxin/Lysophosphatidic Acid/Interleukin-6 Amplification Loop Drives Scleroderma Fibrosis." *Arthritis & Rheumatology* 68(12): 2964-2974.
63. Cecchini, M. J., K. Hosein, C. J. Howlett, M. Joseph and M. Mura (2018). "Comprehensive gene expression profiling identifies distinct and overlapping transcriptional profiles in non-specific interstitial pneumonia and idiopathic pulmonary fibrosis." *Respiratory Research* 19(1): 153.
64. Chanda, D., A. Kurundkar, S. Rangarajan, M. Locy, K. Bernard, N. S. Sharma, N. J. Logsdon, H. Liu, D. K. Crossman, J. C. Horowitz, S. De Langhe and V. J. Thannickal (2016). "Developmental

- Reprogramming in Mesenchymal Stromal Cells of Human Subjects with Idiopathic Pulmonary Fibrosis." *Scientific Reports* 6(1): 37445.
65. Chandriani, S., D. J. DePianto, E. N. N'Diaye, A. R. Abbas, J. Jackman, J. Bevers, V. Ramirez-Carrozzi, R. Pappu, S. E. Kauder, K. Toy, C. Ha, Z. Modrusan, L. C. Wu, H. R. Collard, P. J. Wolters, J. G. Egen and J. R. Arron (2014). "Endogenously Expressed IL-13R α 2 Attenuates IL-13-Mediated Responses but Does Not Activate Signaling in Human Lung Fibroblasts." *The Journal of Immunology* 193(1): 111.
 66. Charbonneau, M., R. R. Lavoie, A. Lauzier, K. Harper, P. P. McDonald and C. M. Dubois (2016). "Platelet-Derived Growth Factor Receptor Activation Promotes the Prodestructive Invadosome-Forming Phenotype of Synoviocytes from Patients with Rheumatoid Arthritis." *The Journal of Immunology* 196(8): 3264-3275.
 67. Cheishvili, D., B. Stefanska, C. Yi, C. Chen Li, P. Yu, A. Arakelian, I. Tanvir, H. Ahmed Khan, S. Rabbani and M. Szyf (2015). "A common promoter hypomethylation signature in invasive breast, liver and prostate cancer cell lines reveals novel targets involved in cancer invasiveness." *Oncotarget*; Vol 6, No 32.
 68. Chen, E. Y., C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark and A. Ma'ayan (2013). "Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool." *BMC Bioinformatics* 14(1): 128.
 69. Chen, G., T. G. Gharib, C.-C. Huang, J. M. G. Taylor, D. E. Misek, S. L. R. Kardia, T. J. Giordano, M. D. Iannettoni, M. B. Orringer, S. M. Hanash and D. G. Beer (2002). "Discordant Protein and mRNA Expression in Lung Adenocarcinomas *." *Molecular & Cellular Proteomics* 1(4): 304-313.
 70. Chen, G., J. C. Ramírez, N. Deng, X. Qiu, C. Wu, W. J. Zheng and H. Wu (2019). "Restructured GEO: Restructuring Gene Expression Omnibus metadata for genome dynamics analysis." *Database* 2019: 1-8.
 71. Chen, M. and K. L. O'Connor (2005). "Integrin α 6 β 4 promotes expression of autotaxin/ENPP2 autocrine motility factor in breast carcinoma cells." *Oncogene* 24(32): 5125-5130.
 72. Chen, R. and J. Dai (2023). "Lipid metabolism in idiopathic pulmonary fibrosis: From pathogenesis to therapy." *Journal of Molecular Medicine* 101(8): 905-915.
 73. Chen, T. and C. Guestrin (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, California, USA, Association for Computing Machinery: 785–794.
 74. Chen, Y. and X. Wang (2020). "MiRDB: An online database for prediction of functional microRNA targets." *Nucleic Acids Research* 48(D1): D127-D131.
 75. Cho, I. and M. J. Blaser (2012). "The human microbiome: at the interface of health and disease." *Nature Reviews Genetics* 13(4): 260-270.
 76. Cho, J.-H., R. Gelinas, K. Wang, A. Etheridge, M. G. Piper, K. Batte, D. Dakhllallah, J. Price, D. Bornman, S. Zhang, C. Marsh and D. Galas (2011). "Systems biology of interstitial lung diseases: integration of mRNA and microRNA expression changes." *BMC Medical Genomics* 4(1): 8.
 77. Choi, J., J.-E. Park, G. Tsagkogeorga, M. Yanagita, B.-K. Koo, N. Han and J.-H. Lee (2020). "Inflammatory Signals Induce AT2 Cell-Derived Damage-Associated Transient Progenitors that Mediate Alveolar Regeneration." *Cell Stem Cell* 27(3): 366-382.e367.
 78. Choi, Y., T. T. Liu, D. G. Pankratz, T. V. Colby, N. M. Barth, D. A. Lynch, P. S. Walsh, G. Raghu, G. C. Kennedy and J. Huang (2018). "Identification of usual interstitial pneumonia pattern using RNA-Seq and machine learning: challenges and solutions." *BMC Genomics* 19(2): 101.
 79. Choi, Y., J. Lu, Z. Hu, D. G. Pankratz, H. Jiang, M. Cao, C. Marchisano, J. Huiras, G. Fedorowicz, M. G. Wong, J. R. Anderson, E. Y. Tom, J. Babiarz, U. Imtiaz, N. M. Barth, P. S. Walsh, G. C. Kennedy and J. Huang (2017). "Analytical performance of Envisia: a genomic classifier for usual interstitial pneumonia." *BMC Pulmonary Medicine* 17(1): 141.
 80. Cisse, B., M. L. Caton, M. Lehner, T. Maeda, S. Scheu, R. Locksley, D. Holmberg, C. Zweier, N. S. den Hollander, S. G. Kant, W. Holter, A. Rauch, Y. Zhuang and B. Reizis (2008). "Transcription

- Factor E2-2 Is an Essential and Specific Regulator of Plasmacytoid Dendritic Cell Development." *Cell* 135(1): 37-48.
81. Colaprico, A., T. C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T. S. Sabedot, T. M. Malta, S. M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi and H. Noushmehr (2016). "TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data." *Nucleic Acids Research* 44(8): e71-e71.
 82. Condamine, T., L. Le Texier, D. Howie, A. Lavault, M. Hill, F. Halary, S. Cobbold, H. Waldmann, M.-C. Cuturi and E. Chiffolleau (2010). "Tmem176B and Tmem176A are associated with the immature state of dendritic cells." *Journal of Leukocyte Biology* 88(3): 507-515.
 83. Conesa, A. and S. Beck (2019). "Making multi-omics data accessible to researchers." *Scientific Data* 6(1): 251.
 84. The Gene Ontology Consortium (2019). "The Gene Ontology Resource: 20 years and still GOing strong." *Nucleic acids research* 47(D1): D330-D338.
 85. The UniProt Consortium (2018). "UniProt: a worldwide hub of protein knowledge." *Nucleic Acids Research* 47(D1): D506-D515.
 86. Cope, L. M., R. A. Irizarry, H. A. Jaffee, Z. Wu and T. P. Speed (2004). "A benchmark for Affymetrix GeneChip expression measures." *Bioinformatics* 20(3): 323-331.
 87. Cordero, F., M. Botta and R. A. Calogero (2007). "Microarray data analysis and mining approaches." *Briefings in Functional Genomics* 6(4): 265-281.
 88. Cosgrove, G. P., K. K. Brown, W. P. Schiemann, A. E. Serls, J. E. Parr, M. W. Geraci, M. I. Schwarz, C. D. Cool and G. S. Worthen (2004). "Pigment Epithelium-derived Factor in Idiopathic Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 170(3): 242-251.
 89. Cosgrove, G. P., M. I. Schwarz, M. W. Geraci, K. K. Brown and G. S. Worthen (2002). "Overexpression of Matrix Metalloproteinase-7 in Pulmonary Fibrosis." *CHEST* 121(3): 25S-26S.
 90. Costa, F. F. (2014). "Big data in biomedicine." *Drug Discovery Today* 19(4): 433-440.
 91. Coward, W. R., K. Watts, C. A. Feghali-Bostwick, A. Knox and L. Pang (2009). "Defective Histone Acetylation Is Responsible for the Diminished Expression of Cyclooxygenase 2 in Idiopathic Pulmonary Fibrosis." *Molecular and Cellular Biology* 29(15): 4325-4339.
 92. Csárdi, G. and T. Nepusz (2006). The igraph software package for complex network research.
 93. Cui, Y., J. Ji, J. Hou, Y. Tan and X. Han (2021). "Identification of Key Candidate Genes Involved in the Progression of Idiopathic Pulmonary Fibrosis." *Molecules* 26(4).
 94. Cummings, R., Y. Zhao, D. Jacoby, E. W. Spannake, M. Ohba, J. G. N. Garcia, T. Watkins, D. He, B. Saatian and V. Natarajan (2004). "Protein Kinase C β ; Mediates Lysophosphatidic Acid-induced NF- κ B Activation and Interleukin-8 Secretion in Human Bronchial Epithelial Cells *." *Journal of Biological Chemistry* 279(39): 41085-41094.
 95. D'Alessandro-Gabazza, C. N., T. Kobayashi, T. Yasuma, M. Toda, H. Kim, H. Fujimoto, O. Hataji, A. Takeshita, K. Nishihama, T. Okano, Y. Okano, Y. Nishii, A. Tomaru, K. Fujiwara, V. F. D'Alessandro, A. M. Abdel-Hamid, Y. Ren, G. V. Pereira, C. L. Wright, A. Hernandez, C. J. Fields, P. M. Yau, S. Wang, A. Mizoguchi, M. Fukumura, J. Ohtsuka, T. Nosaka, K. Kataoka, Y. Kondoh, J. Wu, H. Kawagishi, Y. Yano, R. I. Mackie, I. Cann and E. C. Gabazza (2020). "A Staphylococcus pro-apoptotic peptide induces acute exacerbation of pulmonary fibrosis." *Nature Communications* 11(1): 1539.
 96. Dackor, R. T., J. Cheng, J. W. Voltz, J. W. Card, C. D. Ferguson, R. C. Garrett, J. A. Bradbury, L. M. DeGraff, F. B. Lih, K. B. Tomer, G. P. Flake, G. S. Travlos, R. W. Ramsey, M. L. Edin, D. L. Morgan and D. C. Zeldin (2011). "Prostaglandin E2 protects murine lungs from bleomycin-induced pulmonary fibrosis and lung dysfunction." *American Journal of Physiology-Lung Cellular and Molecular Physiology* 301(5): L645-L655.
 97. Davis, S. and P. S. Meltzer (2007). "GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor." *Bioinformatics* 23(14): 1846-1847.

98. de Zeeuw, D., P. Bekker, E. Henkel, C. Hasslacher, I. Gouni-Berthold, H. Mehling, A. Potarca, V. Tesar, H. J. L. Heerspink and T. J. Schall (2015). "The effect of CCR2 inhibitor CCX140-B on residual albuminuria in patients with type 2 diabetes and nephropathy: a randomised trial." *The Lancet Diabetes & Endocrinology* 3(9): 687-696.
99. Deng, X., M. Xu, C. Yuan, L. Yin, X. Chen, X. Zhou, G. Li, Y. Fu, C. A. Feghali-Bostwick and L. Pang (2013). "Transcriptional regulation of increased CCL2 expression in pulmonary fibrosis involves nuclear factor- κ B and activator protein-1." *The International Journal of Biochemistry & Cell Biology* 45(7): 1366-1376.
100. DePianto, D. J., S. Chandriani, A. R. Abbas, G. Jia, E. N. Diaye, P. Caplazi, S. E. Kauder, S. Biswas, S. K. Karnik, C. Ha, Z. Modrusan, M. A. Matthay, J. Kukreja, H. R. Collard, J. G. Egen, P. J. Wolters and J. R. Arron (2015). "Heterogeneous gene expression signatures correspond to distinct lung pathologies and biomarkers of disease severity in idiopathic pulmonary fibrosis." *Thorax* 70(1): 48.
101. DePianto, D. J., J. A. Vander Heiden, K. B. Morshead, K.-H. Sun, Z. Modrusan, G. Teng, P. J. Wolters and J. R. Arron (2021). "Molecular mapping of interstitial lung disease reveals a phenotypically distinct senescent basal epithelial cell population." *JCI Insight*.
102. Desai, O., J. Winkler, M. Minasyan and E. L. Herzog (2018). "The Role of Immune and Inflammatory Cells in Idiopathic Pulmonary Fibrosis." *Frontiers in Medicine* 5.
103. Devireddy, L. R., C. Gazin, X. Zhu and M. R. Green (2005). "A Cell-Surface Receptor for Lipocalin 24p3 Selectively Mediates Apoptosis and Iron Uptake." *Cell* 123(7): 1293-1305.
104. Dimopoulos, A. C., K. Koukoutegos, F. E. Psomopoulos and P. Moulos (2021). "Combining Multiple RNA-Seq Data Analysis Algorithms Using Machine Learning Improves Differential Isoform Expression Analysis." *Methods and Protocols* 4(4).
105. Distler, J. H. W., A.-H. Györfi, M. Ramanujam, M. L. Whitfield, M. Königshoff and R. Lafyatis (2019). "Shared and distinct mechanisms of fibrosis." *Nature Reviews Rheumatology* 15(12): 705-730.
106. Drake, T. M., A. B. Docherty, E. M. Harrison, J. K. Quint, H. Adamali, S. Agnew, S. Babu, C. M. Barber, S. Barratt, E. Bendstrup, S. Bianchi, D. C. Villegas, N. Chaudhuri, F. Chua, R. Coker, W. Chang, A. Crawshaw, L. E. Crowley, D. Dosanjh, C. A. Fiddler, I. A. Forrest, P. M. George, M. A. Gibbons, K. Groom, S. Haney, S. P. Hart, E. Heiden, M. Henry, L.-P. Ho, R. K. Hoyles, J. Hutchinson, K. Hurley, M. Jones, S. Jones, M. Kokosi, M. Kreuter, L. S. MacKay, S. Mahendran, G. Margaritopoulos, M. Molina-Molina, P. L. Molyneaux, A. O'Brien, K. O'Reilly, A. Packham, H. Parfrey, V. Poletti, J. C. Porter, E. Renzoni, P. Rivera-Ortega, A.-M. Russell, G. Saini, L. G. Spencer, G. M. Stella, H. Stone, S. Sturney, D. Thickett, M. Thillai, T. Wallis, K. Ward, A. U. Wells, A. West, M. Wickremasinghe, F. Woodhead, G. Hearson, L. Howard, J. K. Baillie, P. J. M. Openshaw, M. G. Semple, I. Stewart, R. G. Jenkins, J. K. Baillie, M. G. Semple, P. J. M. Openshaw, G. Carson, B. Alex, B. Bach, W. S. Barclay, D. Bogaert, M. Chand, G. S. Cooke, A. B. Docherty, J. Dunning, A. da Silva Filipe, T. Fletcher, C. A. Green, E. M. Harrison, J. A. Hiscox, A. Y. W. Ho, P. W. Horby, S. Ijaz, S. Khoo, P. Klenerman, A. Law, W. S. Lim, A. J. Mentzer, L. Merson, A. M. Meynert, M. Noursadeghi, S. C. Moore, M. Palmarini, W. A. Paxton, G. Pollakis, N. Price, A. Rambaut, D. L. Robertson, C. D. Russell, V. Sancho-Shimizu, J. T. Scott, T. de Silva, L. Sigfrid, T. Solomon, S. Sriskandan, D. Stuart, C. Summers, R. S. Tedder, E. C. Thomson, A. A. R. Thompson, R. S. Thwaites, L. C. W. Turtle, M. Zambon, H. Hardwick, C. Donohue, R. Lyons, F. Griffiths, W. Oosthuyzen, L. Norman, R. Pius, T. M. Drake, C. J. Fairfield, S. Knight, K. A. McLean, D. Murphy, C. A. Shaw, J. Dalton, M. Girvan, E. Saviciute, S. Roberts, J. Harrison, L. Marsh, M. Connor, S. Halpin, C. Jackson, C. Gamble, G. Leeming, A. Law, M. Wham, S. Clohisey, R. Hendry, J. Scott-Brown, W. Greenhalf, V. Shaw, S. McDonald, S. Keating, K. A. Ahmed, J. A. Armstrong, M. Ashworth, I. G. Asiimwe, S. Bakshi, S. L. Barlow, L. Booth, B. Brennan, K. Bullock, B. W. A. Catterall, J. J. Clark, E. A. Clarke, S. Cole, L. Cooper, H. Cox, C. Davis, O. Dincarslan, C. Dunn, P. Dyer, A. Elliott, A. Evans, L. Finch, L. W. S. Fisher, T. Foster, I. Garcia-Dorival, W. Greenhalf, P. Gunning, C. Hartley, A. Ho, R. L. Jensen, C. B. Jones, T. R. Jones, S. Khandaker, K. King, R. T. Kiy, C. Koukorava, A. Lake, S. Lant, D. Latawiec, L. Lavelle-Langham, D. Lefteri, L. Lett, L. A.

- Livoti, M. Mancini, S. McDonald, L. McEvoy, J. McLauchlan, S. Metelmann, N. S. Miah, J. Middleton, J. Mitchell, S. C. Moore, E. G. Murphy, R. Penrice-Randal, J. Pilgrim, T. Prince, W. Reynolds, P. M. Ridley, D. Sales, V. E. Shaw, R. K. Shears, B. Small, K. S. Subramaniam, A. Szemiel, A. Taggart, J. Tanianis-Hughes, J. Thomas, E. Trochu, L. van Tonder, E. Wilcock, J. E. Zhang, K. Adeniji, D. Agranoff, K. Agwuh, D. Ail, A. Alegria, B. Angus, A. Ashish, D. Atkinson, S. Bari, G. Barlow, S. Barnass, N. Barrett, C. Bassford, D. Baxter, M. Beadsworth, J. Bernatoniene, J. Berridge, N. Best, P. Bothma, D. Brealey, R. Brittain-Long, N. Bulteel, T. Burden, A. Burtenshaw, V. Caruth, D. Chadwick, D. Chamblor, N. Chee, J. Child, S. Chukkambotla, T. Clark, P. Collini, C. Cosgrove, J. Cupitt, M.-T. Cutino-Moguel, P. Dark, C. Dawson, S. Dervisevic, P. Donnison, S. Douthwaite, I. DuRand, A. Dushianthan, T. Dyer, C. Evans, C. Eziefula, C. Fegan, A. Finn, D. Fullerton, S. Garg, S. Garg, A. Garg, E. Gkrania-Klotsas, J. Godden, A. Goldsmith, C. Graham, E. Hardy, S. Hartshorn, D. Harvey, P. Havalda, D. B. Hawcutt, M. Hobrok, L. Hodgson, A. Holme, A. Hormis, M. Jacobs, S. Jain, P. Jennings, A. Kaliappan, V. Kasipandian, S. Kegg, M. Kelsey, J. Kendall, C. Kerrison, I. Kerlake, O. Koch, G. Koduri, G. Koshy, S. Laha, S. Laird, S. Larkin, T. Leiner, P. Lillie, J. Limb, V. Linnett, J. Little, M. MacMahon, E. MacNaughton, R. Mankregod, H. Masson, E. Matovu, K. McCullough, R. McEwen, M. Meda, G. Mills, J. Minton, M. Mirfenderesky, K. Mohandas, Q. Mok, J. Moon, E. Moore, P. Morgan, C. Morris, K. Mortimore, S. Moses, M. Mpenge, R. Mulla, M. Murphy, M. Nagel, T. Nagarajan, M. Nelson, I. Otahal, M. Pais, S. Panchatsharam, H. Paraiso, B. Patel, J. Pepperell, M. Peters, M. Phull, S. Pintus, J. S. Pooni, F. Post, D. Price, R. Prout, N. Rae, H. Reschreiter, T. Reynolds, N. Richardson, M. Roberts, D. Roberts, A. Rose, G. Rousseau, B. Ryan, T. Saluja, A. Shah, P. Shanmuga, A. Sharma, A. Shawcross, J. Sizer, M. Shankar-Hari, R. Smith, C. Snelson, N. Spittle, N. Staines, T. Stambach, R. Stewart, P. Subudhi, T. Szakmany, K. Tatham, J. Thomas, C. Thompson, R. Thompson, A. Tridente, D. Tupper-Carey, M. Twagira, A. Ustianowski, N. Vallotton, L. Vincent-Smith, S. Visuvanathan, A. Vuylsteke, S. Waddy, R. Wake, A. Walden, I. Welters, T. Whitehouse, P. Whittaker, A. Whittington, M. Wijesinghe, M. Williams, L. Wilson, S. Wilson, S. Winchester, M. Wiselka, A. Wolverson, D. G. Wooton, A. Workman, B. Yates and P. Young (2020). "Outcome of Hospitalization for COVID-19 in Patients with Interstitial Lung Disease. An International Multicenter Study." *American Journal of Respiratory and Critical Care Medicine* 202(12): 1656-1665.
107. Durinck, S., P. T. Spellman, E. Birney and W. Huber (2009). "Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt." *Nature Protocols* 4(8): 1184-1191.
108. Dyer, E. L. and K. Kording (2023). "Why the simplest explanation isn't always the best." *Proceedings of the National Academy of Sciences* 120(52): e2319169120.
109. Eagan, T. M., J. K. Damás, T. Ueland, M. Voll-Aanerud, T. E. Mollnes, J. A. Hardie, P. S. Bakke and P. Aukrust (2010). "Neutrophil Gelatinase-Associated Lipocalin: A Biomarker in COPD." *CHEST* 138(4): 888-895.
110. Eliopoulos, A. G., C. D. Dumitru, C. C. Wang, J. Cho and P. N. Tschlis (2002). "Induction of COX-2 by LPS in macrophages is regulated by Tpl2-dependent CREB activation signals." *The EMBO Journal* 21(18): 4831-4840-4840.
111. Emblom-Callahan, M. C., M. K. Chhina, O. A. Shlobin, S. Ahmad, E. S. Reese, E. P. R. Iyer, D. N. Cox, R. Brenner, N. A. Burton, G. M. Grant and S. D. Nathan (2010). "Genomic phenotype of non-cultured pulmonary fibroblasts in idiopathic pulmonary fibrosis." *Genomics* 96(3): 134-145.
112. Emo, J., N. Meednu, T. J. Chapman, F. Rezaee, M. Balys, T. Randall, T. Rangasamy and S. N. Georas (2012). "Lpa2 Is a Negative Regulator of Both Dendritic Cell Activation and Murine Models of Allergic Lung Inflammation." *The Journal of Immunology* 188(8): 3784-3790.
113. Espindola, M. S., D. M. Habel, A. L. Coelho, B. Stripp, W. C. Parks, J. Oldham, F. J. Martinez, I. Noth, D. Lopez, A. Mikels-Vigdal, V. Smith and C. M. Hogaboam (2020). "Differential Responses to Targeting Matrix Metalloproteinase 9 in Idiopathic Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 203(4): 458-470.
114. Esposito, A. J., A. A. Menon, A. J. Ghosh, R. K. Putman, L. E. Fredenburgh, S. Y. El-Chemaly, H. J. Goldberg, R. M. Baron, G. M. Hunninghake and T. J. Doyle (2020). "Increased Odds of Death

- for Patients with Interstitial Lung Disease and COVID-19: A Case–Control Study." *American Journal of Respiratory and Critical Care Medicine* 202(12): 1710-1713.
115. Failla, M., T. Genovese, E. Mazzon, M. Fruciano, E. Fagone, E. Gili, A. Barera, C. La Rosa, E. Conte, N. Crimi, S. Cuzzocrea and C. Vancheri (2009). "16,16-Dimethyl Prostaglandin E2 Efficacy on Prevention and Protection from Bleomycin-Induced Lung Injury and Fibrosis." *American Journal of Respiratory Cell and Molecular Biology* 41(1): 50-58.
 116. Fan, Y. and O. Pedersen (2021). "Gut microbiota in human metabolic health and disease." *Nature Reviews Microbiology* 19(1): 55-71.
 117. Fanidis, D. and P. Moulos (2020). "Integrative, normalization-insusceptible statistical analysis of RNA-Seq data, with improved differential expression and unbiased downstream functional analysis." *Briefings in Bioinformatics*.
 118. Fanidis, D., P. Moulos and V. Aidinis (2021). "Fibromine is a multi-omics database and mining tool for target discovery in pulmonary fibrosis." *Scientific Reports* 11(1): 21712.
 119. Fanidis, D., V. C. Pezoulas, D. I. Fotiadis and V. Aidinis (2023). "An explainable machine learning-driven proposal of pulmonary fibrosis biomarkers." *Computational and Structural Biotechnology Journal* 21: 2305-2315.
 120. Farina, A. R., L. Cappabianca, P. Ruggeri, N. Di Ianni, M. Ragone, S. Merolle, K. Sano, M. L. Stracke, J. M. Horowitz, A. Gulino and A. R. Mackay (2012). "Constitutive autotaxin transcription by Nmyc-amplified and non-amplified neuroblastoma cells is regulated by a novel AP-1 and SP-mediated mechanism and abrogated by curcumin." *FEBS Letters* 586(20): 3681-3691.
 121. Fernandez, N. F., G. W. Gundersen, A. Rahman, M. L. Grimes, K. Rikova, P. Hornbeck and A. Ma'ayan (2017). "Clustergrammer, a web-based heatmap visualization and analysis tool for high-dimensional biological data." *Scientific Data* 4(1): 170151.
 122. Fichtner-Feigl, S., W. Strober, K. Kawakami, R. K. Puri and A. Kitani (2006). "IL-13 signaling through the IL-13 α 2 receptor is involved in induction of TGF- β 1 production and fibrosis." *Nature Medicine* 12(1): 99-106.
 123. Fingerlin, T. E., E. Murphy, W. Zhang, A. L. Peljto, K. K. Brown, M. P. Steele, J. E. Loyd, G. P. Cosgrove, D. Lynch, S. Groshong, H. R. Collard, P. J. Wolters, W. Z. Bradford, K. Kossen, S. D. Seiwert, R. M. du Bois, C. K. Garcia, M. S. Devine, G. Gudmundsson, H. J. Isaksson, N. Kaminski, Y. Zhang, K. F. Gibson, L. H. Lancaster, J. D. Cogan, W. R. Mason, T. M. Maher, P. L. Molyneaux, A. U. Wells, M. F. Moffatt, M. Selman, A. Pardo, D. S. Kim, J. D. Crapo, B. J. Make, E. A. Regan, D. S. Walek, J. J. Daniel, Y. Kamatani, D. Zelenika, K. Smith, D. McKean, B. S. Pedersen, J. Talbert, R. N. Kidd, C. R. Markin, K. B. Beckman, M. Lathrop, M. I. Schwarz and D. A. Schwartz (2013). "Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis." *Nature Genetics* 45(6): 613-620.
 124. Fischbach, M. A., H. Lin, D. R. Liu and C. T. Walsh (2006). "How pathogenic bacteria evade mammalian sabotage in the battle for iron." *Nature Chemical Biology* 2(3): 132-138.
 125. Flaumenhaft, R., K. Enjyoji and A. A. Schmaier (2022). "Vasculopathy in COVID-19." *Blood* 140(3): 222-235.
 126. Flo, T. H., K. D. Smith, S. Sato, D. J. Rodriguez, M. A. Holmes, R. K. Strong, S. Akira and A. Aderem (2004). "Lipocalin 2 mediates an innate immune response to bacterial infection by sequestering iron." *Nature* 432(7019): 917-921.
 127. Flower, D. R., A. C. T. North and C. E. Sansom (2000). "The lipocalin protein family: structural and sequence overview." *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* 1482(1): 9-24.
 128. Forlino, A. and J. C. Marini (2016). "Osteogenesis imperfecta." *The Lancet* 387(10028): 1657-1671.
 129. Fotopoulou, S., N. Oikonomou, E. Grigorieva, I. Nikitopoulou, T. Paparountas, A. Thanassopoulou, Z. Zhao, Y. Xu, D. L. Kontoyiannis, E. Remboutsika and V. Aidinis (2010). "ATX

- expression and LPA signalling are vital for the development of the nervous system." *Developmental Biology* 339(2): 451-464.
130. Fraser, E., L. Denney, A. Antanaviciute, K. Blirando, C. Vuppusetty, Y. Zheng, E. Repapi, V. Iotchkova, S. Taylor, N. Ashley, V. St Noble, R. Benamore, R. Hoyles, C. Clelland, J. M. D. Rastrick, C. S. Hardman, N. K. Alham, R. E. Rigby, A. Simmons, J. Rehwinkel and L.-P. Ho (2021). "Multi-Modal Characterization of Monocytes in Idiopathic Pulmonary Fibrosis Reveals a Primed Type I Interferon Immune Phenotype." *Frontiers in Immunology* 12.
 131. Fraser, J. D. and T. Proft (2008). "The bacterial superantigen and superantigen-like proteins." *Immunological Reviews* 225(1): 226-243.
 132. Galaris, A., D. Fanidis, E.-A. Stylianaki, V. Harokopos, A.-S. Kalantzi, P. Moulos, A. S. Dimas, P. Hatzis and V. Aidinis (2022). "Obesity Reshapes the Microbial Population Structure along the Gut-Liver-Lung Axis in Mice." *Biomedicines* 10(2).
 133. Galaris, A., D. Fanidis, E. Tsitoura, P. Kanellopoulou, I. Barbayianni, K. Ntatsoulis, K. Touloumi, S. Gramenoudi, T. Karampitsakos, A. Tzouveleki, K. Antoniou and V. Aidinis (2023). "Increased lipocalin-2 expression in pulmonary inflammation and fibrosis." *Frontiers in Medicine* 10.
 134. Gantke, T., S. Sriskantharajah and S. C. Ley (2011). "Regulation and function of TPL-2, an I κ B kinase-regulated MAP kinase kinase kinase." *Cell Research* 21(1): 131-145.
 135. Gao, L., X. Li, H. Wang, Y. Liao, Y. Zhou, K. Wang, J. Hu, M. Cheng, Z. Zeng, T. Wang and F. Wen (2021). "Autotaxin levels in serum and bronchoalveolar lavage fluid are associated with inflammatory and fibrotic biomarkers and the clinical outcome in patients with acute respiratory distress syndrome." *Journal of Intensive Care* 9(1): 44.
 136. Garcia-Alonso, L., C. H. Holland, M. M. Ibrahim, D. Turei and J. Saez-Rodriguez (2019). "Benchmark and integration of resources for the estimation of human transcription factor activities." *Genome Research* 29(8): 1363-1375.
 137. Gaskill, C. F., E. J. Carrier, J. A. Kropski, N. C. Bloodworth, S. Menon, R. F. Foronjy, M. M. Taketo, C. C. Hong, E. D. Austin, J. D. West, A. L. Means, J. E. Loyd, W. D. Merryman, A. R. Hemnes, S. De Langhe, T. S. Blackwell, D. J. Klemm and S. M. Majka (2017). "Disruption of lineage specification in adult pulmonary mesenchymal progenitor cells promotes microvascular dysfunction." *The Journal of Clinical Investigation* 127(6): 2262-2276.
 138. Gautier, L., L. Cope, B. M. Bolstad and R. A. Irizarry (2004). "affy—analysis of Affymetrix GeneChip data at the probe level." *Bioinformatics* 20(3): 307-315.
 139. Geng, H., R. Lan, P. K. Singha, A. Gilchrist, P. H. Weinreb, S. M. Violette, J. M. Weinberg, P. Saikumar and M. A. Venkatachalam (2012). "Lysophosphatidic Acid Increases Proximal Tubule Cell Secretion of Profibrotic Cytokines PDGF-B and CTGF through LPA2- and G α 12/13-mediated Rho and α 5 β 1 Integrin-Dependent Activation of TGF- β 2." *The American Journal of Pathology* 181(4): 1236-1249.
 140. Geng, J., X. Huang, Y. Li, X. Xu, S. Li, D. Jiang, J. Liang, D. Jiang, C. Wang and H. Dai (2015). "Down-regulation of USP13 mediates phenotype transformation of fibroblasts in idiopathic pulmonary fibrosis." *Respiratory Research* 16(1): 124.
 141. Georganas, C., H. Liu, H. Perlman, A. Hoffmann, B. Thimmapaya and R. M. Pope (2000). "Regulation of IL-6 and IL-8 Expression in Rheumatoid Arthritis Synovial Fibroblasts: the Dominant Role for NF- κ B But Not C/EBP β or c-Jun1." *The Journal of Immunology* 165(12): 7199-7206.
 142. Gershoni, M. and S. Pietrokovski (2017). "The landscape of sex-differential transcriptome and its consequent selection in human adults." *BMC Biology* 15(1): 7.
 143. Gharib, S. A., W. A. Altemeier, L. S. Van Winkle, C. G. Plopper, S. Y. Schlesinger, C. A. Buell, R. Brauer, V. Lee, W. C. Parks and P. Chen (2013). "Matrix Metalloproteinase-7 Coordinates Airway Epithelial Injury Response and Differentiation of Ciliated Cells." *American Journal of Respiratory Cell and Molecular Biology* 48(3): 390-396.

144. Giganti, A., M. Rodriguez, B. Fould, N. Moulharat, F. Cogé, P. Chomarat, J.-P. Galizzi, P. Valet, J.-S. Saulnier-Blache, J. A. Boutin and G. Ferry (2008). "Murine and Human Autotaxin α 1, α 2, and α 3 Isoforms: GENE ORGANIZATION, TISSUE DISTRIBUTION, AND BIOCHEMICAL CHARACTERIZATION *." *Journal of Biological Chemistry* 283(12): 7776-7789.</sup>
145. Giurgiu, M., J. Reinhard, B. Brauner, I. Dunger-Kaltenbach, G. Fobo, G. Frishman, C. Montrone and A. Ruepp (2019). "CORUM: the comprehensive resource of mammalian protein complexes—2019." *Nucleic Acids Research* 47(D1): D559-D563.
146. Gkirtzimanaki, K., K. K. Gkouskou, U. Oleksiewicz, G. Nikolaidis, D. Vyrla, M. Lontos, V. Pelekanou, D. C. Kanellis, K. Evangelou, E. N. Stathopoulos, J. K. Field, P. N. Tsihchlis, V. Gorgoulis, T. Liloglou and A. G. Eliopoulos (2013). "TPL2 kinase is a suppressor of lung carcinogenesis." *Proceedings of the National Academy of Sciences* 110(16): E1470-E1479.
147. Gloor, G. B., J. M. Macklaim, V. Pawlowsky-Glahn and J. J. Egozcue (2017). "Microbiome Datasets Are Compositional: And This Is Not Optional." *Frontiers in Microbiology* 8: 2224.
148. Goetz, D. H., M. A. Holmes, N. Borregaard, M. E. Bluhm, K. N. Raymond and R. K. Strong (2002). "The Neutrophil Lipocalin NGAL Is a Bacteriostatic Agent that Interferes with Siderophore-Mediated Iron Acquisition." *Molecular Cell* 10(5): 1033-1043.
149. Goodwin, S., J. D. McPherson and W. R. McCombie (2016). "Coming of age: ten years of next-generation sequencing technologies." *Nature Reviews Genetics* 17(6): 333-351.
150. Gralinski, L. E., M. T. Ferris, D. L. Aylor, A. C. Whitmore, R. Green, M. B. Frieman, D. Deming, V. D. Menachery, D. R. Miller, R. J. Buus, T. A. Bell, G. A. Churchill, D. W. Threadgill, M. G. Katze, L. McMillan, W. Valdar, M. T. Heise, F. Pardo-Manuel de Villena and R. S. Baric (2015). "Genome Wide Identification of SARS-CoV Susceptibility Loci Using the Collaborative Cross." *PLOS Genetics* 11(10): e1005504.
151. Gralinski, L. E., V. D. Menachery, A. P. Morgan, A. L. Tatura, A. Beall, J. Kocher, J. Plante, D. C. Harrison-Shostak, A. Schäfer, F. Pardo-Manuel de Villena, M. T. Ferris and R. S. Baric (2017). "Allelic Variation in the Toll-Like Receptor Adaptor Protein Ticam2 Contributes to SARS-Coronavirus Pathogenesis in Mice." *G3 Genes|Genomes|Genetics* 7(6): 1653-1663.
152. Grynberg, K., F. Y. Ma and D. J. Nikolic-Paterson (2017). "The JNK Signaling Pathway in Renal Fibrosis." *Frontiers in Physiology* 8.
153. Gu, Z., L. Gu, R. Eils, M. Schlesner and B. Brors (2014). "circlize implements and enhances circular visualization in R." *Bioinformatics* 30(19): 2811-2812.
154. Guan, S. and J. Zhou (2017). "Frizzled-7 mediates TGF- β -induced pulmonary fibrosis by transmitting non-canonical Wnt signaling." *Experimental Cell Research* 359(1): 226-234.
155. Guardado, S., D. Ojeda-Juárez, M. Kaul and T. M. Nordgren (2021). "Comprehensive review of lipocalin 2-mediated effects in lung inflammation." *American Journal of Physiology-Lung Cellular and Molecular Physiology* 321(4): L726-L733.
156. Guo, L., P. He, Y. R. No and C. C. Yun (2015). "Krüppel-like factor 5 incorporates into the β -catenin/TCF complex in response to LPA in colon cancer cells." *Cellular Signalling* 27(5): 961-968.
157. Guo, X., C. Sunil and G. Qian (2022). "Obesity and the Development of Lung Fibrosis." *Frontiers in Pharmacology* 12.
158. Habermann, A. C., A. J. Gutierrez, L. T. Bui, S. L. Yahn, N. I. Winters, C. L. Calvi, L. Peter, M.-I. Chung, C. J. Taylor, C. Jetter, L. Raju, J. Roberson, G. Ding, L. Wood, J. M. S. Sucre, B. W. Richmond, A. P. Serezani, W. J. McDonnell, S. B. Mallal, M. J. Bacchetta, J. E. Loyd, C. M. Shaver, L. B. Ware, R. Bremner, R. Walia, T. S. Blackwell, N. E. Banovich and J. A. Kropski (2020). "Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis." *Science Advances* 6(28): eaba1972.
159. Habel, D. M., A. Camelo, M. Espindola, T. Burwell, R. Hanna, E. Miranda, A. Carruthers, M. Bell, A. L. Coelho, H. Liu, F. Pilataxi, L. Clarke, E. Grant, A. Lewis, B. Moore, D. A. Knight, C.

- M. Hogaboam and L. A. Murray (2017). "Divergent roles for Clusterin in Lung Injury and Repair." *Scientific Reports* 7(1): 15444.
160. Haij, S. d., A. M. Woltman, A. C. Bakker, M. R. Daha and C. van Kooten (2002). "Production of inflammatory mediators by renal epithelial cells is insensitive to glucocorticoids." *British Journal of Pharmacology* 137(2): 197-204.
161. Han, M. K., Y. Zhou, S. Murray, N. Tayob, I. Noth, V. N. Lama, B. B. Moore, E. S. White, K. R. Flaherty, G. B. Huffnagle and F. J. Martinez (2014). "Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study." *The Lancet Respiratory Medicine* 2(7): 548-556.
162. Han, N., J. He, L. Shi, M. Zhang, J. Zheng and Y. Fan (2022). "Identification of biomarkers in nonalcoholic fatty liver disease: A machine learning method and experimental study." *Frontiers in Genetics* 13.
163. Hancock, L. A., C. E. Hennessy, G. M. Solomon, E. Dobrinskikh, A. Estrella, N. Hara, D. B. Hill, W. J. Kissner, M. R. Markovetz, D. E. Grove Villalon, M. E. Voss, G. J. Tearney, K. S. Carroll, Y. Shi, M. I. Schwarz, W. R. Thelin, S. M. Rowe, I. V. Yang, C. M. Evans and D. A. Schwartz (2018). "Muc5b overexpression causes mucociliary dysfunction and enhances lung fibrosis in mice." *Nature Communications* 9(1): 5363.
164. Harbeck, N., F. Penault-Llorca, J. Cortes, M. Gnant, N. Houssami, P. Poortmans, K. Ruddy, J. Tsang and F. Cardoso (2019). "Breast cancer." *Nature Reviews Disease Primers* 5(1): 66.
165. Harr, B. and C. Schlötterer (2006). "Comparison of algorithms for the analysis of Affymetrix microarray data as evaluated by co-expression of genes in known operons." *Nucleic Acids Research* 34(2): e8-e8.
166. Hayden, M. S. and S. Ghosh (2012). "NF- κ B, the first quarter-century: remarkable progress and outstanding questions." *Genes & Development* 26(3): 203-234.
167. He, F.-F., D. Zhang, Q. Chen, Y. Zhao, L. Wu, Z.-Q. Li, C. Zhang, Z.-H. Jiang and Y.-M. Wang (2019). "Angiotensin-Tie signaling in kidney diseases: an updated review." *FEBS Letters* 593(19): 2706-2715.
168. Héberlé, É. and Anaïs F. Bardet (2019). "Sensitivity of transcription factors to DNA methylation." *Essays in Biochemistry* 63(6): 727-741.
169. Hecker, L., R. Vittal, T. Jones, R. Jagirdar, T. R. Luckhardt, J. C. Horowitz, S. Pennathur, F. J. Martinez and V. J. Thannickal (2009). "NADPH oxidase-4 mediates myofibroblast activation and fibrogenic responses to lung injury." *Nature Medicine* 15(9): 1077-1081.
170. Helsen, J., J. Frickel, R. Jelier and K. J. Verstrepen (2019). "Network hubs affect evolvability." *PLOS Biology* 17(1): e3000111.
171. Henderson, N. C., F. Rieder and T. A. Wynn (2020). "Fibrosis: from mechanisms to medicines." *Nature* 587(7835): 555-566.
172. Herazo-Maya, J. D., I. Noth, S. R. Duncan, S. Kim, S.-F. Ma, G. C. Tseng, E. Feingold, B. M. Juan-Guardela, T. J. Richards, Y. Lussier, Y. Huang, R. Vij, K. O. Lindell, J. Xue, K. F. Gibson, S. D. Shapiro, J. G. N. Garcia and N. Kaminski (2013). "Peripheral Blood Mononuclear Cell Gene Expression Profiles Predict Poor Outcome in Idiopathic Pulmonary Fibrosis." *Science Translational Medicine* 5(205): 205ra136.
173. Hider, R. C. and X. Kong (2010). "Chemistry and biology of siderophores." *Natural Product Reports* 27(5): 637-657.
174. Hilberg, O., E. Bendstrup, R. Ibsen, A. Løkke and C. Hyldgaard (2018). "Economic consequences of idiopathic pulmonary fibrosis in Denmark." *ERJ Open Research* 4(2): 00045-02017.
175. Holzberg, D., C. G. Knight, O. Dittrich-Breiholz, H. Schneider, A. Dörrie, E. Hoffmann, K. Resch and M. Kracht (2003). "Disruption of the c-JUN-JNK Complex by a Cell-permeable Peptide Containing the c-JUN Δ 3b4; Domain Induces Apoptosis and Affects a Distinct Set of Interleukin-1-induced Inflammatory Genes *." *Journal of Biological Chemistry* 278(41): 40213-40223.

176. Horimasu, Y., N. Ishikawa, M. Taniwaki, K. Yamaguchi, K. Hamai, H. Iwamoto, S. Ohshimo, H. Hamada, N. Hattori, M. Okada, K. Arihiro, Y. Ohtsuki and N. Kohno (2017). "Gene expression profiling of idiopathic interstitial pneumonias (IIPs): identification of potential diagnostic markers and therapeutic targets." *BMC Medical Genetics* 18(1): 88.
177. Hoshino, D., K. M. Branch and A. M. Weaver (2013). "Signaling inputs to invadopodia and podosomes." *Journal of Cell Science* 126(14): 2979-2989.
178. Hsu, E., H. Shi, R. M. Jordan, J. Lyons-Weiler, J. M. Pilewski and C. A. Feghali-Bostwick (2011). "Lung tissues in patients with systemic sclerosis have gene expression patterns unique to pulmonary fibrosis and pulmonary hypertension." *Arthritis & Rheumatism* 63(3): 783-794.
179. Hu, B., H. Guo, P. Zhou and Z.-L. Shi (2021). "Characteristics of SARS-CoV-2 and COVID-19." *Nature Reviews Microbiology* 19(3): 141-154.
180. Hu, C., T. Li, Y. Xu, X. Zhang, F. Li, J. Bai, J. Chen, W. Jiang, K. Yang, Q. Ou, X. Li, P. Wang and Y. Zhang (2023). "CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data." *Nucleic Acids Research* 51(D1): D870-D876.
181. Hu, H., Y. R. Miao, L. H. Jia, Q. Y. Yu, Q. Zhang and A. Y. Guo (2019). "AnimalTFDB 3.0: A comprehensive resource for annotation and prediction of animal transcription factors." *Nucleic Acids Research* 47(D1): D33-D38.
182. Hu, S.-L., C.-C. Huang, T.-T. Tzeng, S.-C. Liu, C.-H. Tsai, Y.-C. Fong and C.-H. Tang (2020). "S1P promotes IL-6 expression in osteoblasts through the PI3K, MEK/ERK and NF- κ B signaling pathways." *International Journal of Medical Sciences* 17(9): 1207-1214.
183. Huan, C., T. Yang, J. Liang, T. Xie, L. Cheng, N. Liu, A. Kurkciyan, J. Monterrosa Mena, C. Wang, H. Dai, P. W. Noble and D. Jiang (2015). "Methylation-mediated BMPER expression in fibroblast activation in vitro and lung fibrosis in mice in vivo." *Scientific Reports* 5(1): 14910.
184. Huang, K. Y. and E. Petretto (2021). "Cross-species integration of single-cell RNA-seq resolved alveolar-epithelial transitional states in idiopathic pulmonary fibrosis." *American Journal of Physiology-Lung Cellular and Molecular Physiology* 321(3): L491-L506.
185. Huang, L. S., E. V. Berdyshev, J. T. Tran, L. Xie, J. Chen, D. L. Ebenezer, B. Mathew, I. Gorshkova, W. Zhang, S. P. Reddy, A. Harijith, G. Wang, C. Feghali-Bostwick, I. Noth, S.-F. Ma, T. Zhou, W. Ma, J. G. N. Garcia and V. Natarajan (2015). "Sphingosine-1-phosphate lyase is an endogenous suppressor of pulmonary fibrosis: role of S1P signalling and autophagy." *Thorax* 70(12): 1138.
186. Huang, L. S., B. Mathew, H. Li, Y. Zhao, S.-F. Ma, I. Noth, S. P. Reddy, A. Harijith, P. V. Usatyuk, E. V. Berdyshev, N. Kaminski, T. Zhou, W. Zhang, Y. Zhang, J. Rehman, S. R. Kotha, T. O. Gurney, N. L. Parinandi, Y. A. Lussier, J. G. N. Garcia and V. Natarajan (2014). "The Mitochondrial Cardiolipin Remodeling Enzyme Lysocardiolipin Acyltransferase Is a Novel Target in Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 189(11): 1402-1415.
187. Huang, S., N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang and W. Xu (2018). "Applications of Support Vector Machine (SVM) Learning in Cancer Genomics." *Cancer Genomics - Proteomics* 15(1): 41.
188. Huang, S., F. Lee, R. Miao, Q. Si, C. Lu and Q. Chen (2020). "A deep convolutional neural network architecture for interstitial lung disease pattern classification." *Medical & Biological Engineering & Computing* 58(4): 725-737.
189. Huang, Y., S.-F. Ma, M. S. Espindola, R. Vij, J. M. Oldham, G. B. Huffnagle, J. R. Erb-Downward, K. R. Flaherty, B. B. Moore, E. S. White, T. Zhou, J. Li, Y. A. Lussier, M. K. Han, N. Kaminski, J. G. N. Garcia, C. M. Hogaboam, F. J. Martinez and I. Noth (2017). "Microbes Are Associated with Host Innate Immune Response in Idiopathic Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 196(2): 208-219.

190. Huang, Z., H. Li, S. Liu, J. Jia, Y. Zheng and B. Cao (2022). "Identification of Neutrophil-Related Factor LCN2 for Predicting Severity of Patients With Influenza A Virus and SARS-CoV-2 Infection." *Frontiers in Microbiology* 13.
191. Huleihel, L., A. Ben-Yehudah, J. Milosevic, G. Yu, K. Pandit, K. Sakamoto, H. Yousef, M. LeJeune, T. A. Coon, C. J. Redinger, L. Chensny, E. Manor, G. Schatten and N. Kaminski (2014). "Let-7d microRNA affects mesenchymal phenotypic properties of lung fibroblasts." *American Journal of Physiology-Lung Cellular and Molecular Physiology* 306(6): L534-L542.
192. Hutchinson, J. P., T. M. McKeever, A. W. Fogarty, V. Navaratnam and R. B. Hubbard (2014). "Increasing Global Mortality from Idiopathic Pulmonary Fibrosis in the Twenty-First Century." *Annals of the American Thoracic Society* 11(8): 1176-1185.
193. Hwang, Y. S., M. Jeong, J. S. Park, M. H. Kim, D. B. Lee, B. A. Shin, N. Mukaida, L. M. Ellis, H. R. Kim, B. W. Ahn and Y. D. Jung (2004). "Interleukin-1 β stimulates IL-8 expression through MAP kinase and ROS signaling in human gastric carcinoma cells." *Oncogene* 23(39): 6603-6611.
194. Iizuka, S., R. P. Leon, K. P. Gribbin, Y. Zhang, J. Navarro, R. Smith, K. Devlin, L. G. Wang, S. L. Gibbs, J. Korkola, X. Nan and S. A. Courtneidge (2020). "Crosstalk between invadopodia and the extracellular matrix." *European Journal of Cell Biology* 99(7): 151122.
195. Ikezoe, K., T. Handa, K. Mori, K. Watanabe, K. Tanizawa, K. Aihara, T. Tsuruyama, A. Miyagawa-Hayashino, A. Sokai, T. Kubo, S. Muro, S. Nagai, T. Hirai, K. Chin and M. Mishima (2014). "Neutrophil gelatinase-associated lipocalin in idiopathic pulmonary fibrosis." *European Respiratory Journal* 43(6): 1807.
196. Irizarry, R. A., B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf and T. P. Speed (2003). "Exploration, normalization, and summaries of high density oligonucleotide array probe level data." *Biostatistics* 4(2): 249-264.
197. Isshiki, T., S. Sakamoto, A. Yamasaki, H. Shimizu, S. Miyoshi, Y. Nakamura, S. Homma and K. Kishi (2021). "Incidence of acute exacerbation of idiopathic pulmonary fibrosis in patients receiving antifibrotic agents: Real-world experience." *Respiratory Medicine* 187.
198. Ivan, J., G. Patricia and D. Agustriawan (2021). "In silico study of cancer stage-specific DNA methylation pattern in White breast cancer patients based on TCGA dataset." *Computational Biology and Chemistry* 92: 107498.
199. Jaber, S. A., A. Cohen, C. D'Souza, Y. M. Abdulrazzaq, S. Ojha, S. Bastaki and E. A. Adeghate (2021). "Lipocalin-2: Structure, function, distribution and role in metabolic disorders." *Biomedicine & Pharmacotherapy* 142: 112002.
200. Jaeger, B., J. C. Schupp, L. Plappert, O. Terwolbeck, N. Artysh, G. Kayser, P. Engelhard, T. S. Adams, R. Zweigerdt, H. Kempf, S. Lienenklaus, W. Garrels, I. Nazarenko, D. Jonigk, M. Wygrecka, D. Klatt, A. Schambach, N. Kaminski and A. Prasse (2022). "Airway basal cells show a dedifferentiated KRT17^{high} Phenotype and promote fibrosis in idiopathic pulmonary fibrosis." *Nature Communications* 13(1): 5637.
201. Jang, E., S. Lee, J.-H. Kim, J.-H. Kim, J.-W. Seo, W.-H. Lee, K. Mori, K. Nakao and K. Suk (2013). "Secreted protein lipocalin-2 promotes microglial M1 polarization." *The FASEB Journal* 27(3): 1176-1190.
202. Jiang, A., N. Liu, J. Wang, X. Zheng, M. Ren, W. Zhang and Y. Yao (2022). "The role of PD-1/PD-L1 axis in idiopathic pulmonary fibrosis: Friend or foe?" *Frontiers in Immunology* 13.
203. Jiang, P., R. Gil de Rubio, S. M. Hrycaj, S. J. Gurczynski, K. A. Riemondy, B. B. Moore, M. B. Omary, K. M. Ridge and R. L. Zemans (2020). "Ineffectual Type 2-to-Type 1 Alveolar Epithelial Cell Differentiation in Idiopathic Pulmonary Fibrosis: Persistence of the KRT8^{hi} Transitional State." *American Journal of Respiratory and Critical Care Medicine* 201(11): 1443-1447.
204. Jiang, W., N. Wu, X. Wang, Y. Chi, Y. Zhang, X. Qiu, Y. Hu, J. Li and Y. Liu (2015). "Dysbiosis gut microbiota associated with inflammation and impaired mucosal immune function in intestine of humans with non-alcoholic fatty liver disease." *Scientific Reports* 5(1): 8096.

205. Johnson, G. L., T. J. Stuhlmiller, S. P. Angus, J. S. Zawistowski and L. M. Graves (2014). "Molecular Pathways: Adaptive Kinome Reprogramming in Response to Targeted Inhibition of the BRAF–MEK–ERK Pathway in Cancer." *Clinical Cancer Research* 20(10): 2516-2522.
206. Johnson, J. S., D. J. Spakowicz, B.-Y. Hong, L. M. Petersen, P. Demkowicz, L. Chen, S. R. Leopold, B. M. Hanson, H. O. Agresta, M. Gerstein, E. Sodergren and G. M. Weinstock (2019). "Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis." *Nature Communications* 10(1): 5029.
207. Jolliffe, I. T. and J. Cadima (2016). "Principal component analysis: a review and recent developments." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374(2065): 20150202.
208. Jordan, M. I. and T. M. Mitchell (2015). "Machine learning: Trends, perspectives, and prospects." *Science* 349(6245): 255-260.
209. Joshi, N., S. Watanabe, R. Verma, R. P. Jablonski, C.-I. Chen, P. Cheresch, N. S. Markov, P. A. Reyfman, A. C. McQuattie-Pimentel, L. Sichizya, Z. Lu, R. Piseaux, D. Kirchenbuechler, A. S. Flozak, C. J. Gottardi, C. M. Cuda, H. Perlman, M. Jain, D. W. Kamp, G. R. Scott Budinger and A. V. Misharin (2019). "A spatially restricted fibrotic niche in pulmonary fibrosis is sustained by M-CSF/M-CSFR signaling in monocyte-derived alveolar macrophages." *European Respiratory Journal*: 1900646.
210. Kabuyama, Y., K. Oshima, T. Kitamura, M. Homma, J. Yamaki, M. Munakata and Y. Homma (2007). "Involvement of selenoprotein P in the regulation of redox balance and myofibroblast viability in idiopathic pulmonary fibrosis." *Genes to Cells* 12(11): 1235-1244.
211. Kaminski, N., J. D. Allard, J. F. Pittet, F. Zuo, M. J. D. Griffiths, D. Morris, X. Huang, D. Sheppard and R. A. Heller (2000). "Global analysis of gene expression in pulmonary fibrosis reveals distinct programs regulating lung inflammation and fibrosis." *Proceedings of the National Academy of Sciences* 97(4): 1778-1783.
212. Kan, M., M. Shumyatcher and B. E. Himes (2017). "Using omics approaches to understand pulmonary diseases." *Respiratory Research* 18(1): 149.
213. Kanehisa, M., Y. Sato, M. Furumichi, K. Morishima and M. Tanabe (2019). "New approach for understanding genome variations in KEGG." *Nucleic Acids Research* 47(D1): D590-D595.
214. Kang, S. S., Y. Ren, C. C. Liu, A. Kurti, K. E. Baker, G. Bu, Y. Asmann and J. D. Fryer (2018). "Lipocalin-2 protects the brain during inflammatory conditions." *Molecular Psychiatry* 23(2): 344-350.
215. Kangelaris, K. N., A. Prakash, K. D. Liu, B. Aouizerat, P. G. Woodruff, D. J. Erle, A. Rogers, E. J. Seeley, J. Chu, T. Liu, T. Osterberg-Deiss, H. Zhuo, M. A. Matthay and C. S. Calfee (2015). "Increased expression of neutrophil-related genes in patients with early sepsis-induced ARDS." *American Journal of Physiology-Lung Cellular and Molecular Physiology* 308(11): L1102-L1113.
216. Kaniaris, E., K. Vaporidi, E. Vergadi, E. E. Theodorakis, E. Kondili, E. Lagoudaki, C. Tsatsanis and D. Georgopoulos (2014). "Genetic and pharmacologic inhibition of Tpl2 kinase is protective in a mouse model of ventilator-induced lung injury." *Intensive Care Medicine Experimental* 2(1): 15.
217. Karampitsakos, T., V. Tzilas, R. Tringidou, P. Steiropoulos, V. Aidinis, S. A. Papis, D. Bouros and A. Tzouveleki (2017). "Lung cancer in patients with idiopathic pulmonary fibrosis." *Pulmonary Pharmacology & Therapeutics* 45: 1-10.
218. Karatzas, E., M. M. Bourdakou, G. Kolios and G. M. Spyrou (2017). "Drug repurposing in idiopathic pulmonary fibrosis filtered by a bioinformatics-derived composite score." *Scientific Reports* 7(1): 12569.
219. Karvonen, H. M., S. T. Lehtonen, R. T. Sormunen, T. H. Harju, E. Lappi-Blanco, R. S. Bloigu and R. L. Kaarteenaho (2012). "Myofibroblasts in interstitial lung diseases show diverse electron microscopic and invasive features." *Laboratory Investigation* 92(9): 1270-1284.
220. Kass, D. J., G. Yu, K. S. Loh, A. Savir, A. Borczuk, R. Kahloon, B. Juan-Guardela, G. Deiullis, J. Tedrow, J. Choi, T. Richards, N. Kaminski and S. M. Greenberg (2012). "Cytokine-Like Factor 1

- Gene Expression Is Enriched in Idiopathic Pulmonary Fibrosis and Drives the Accumulation of CD4⁺ T Cells in Murine Lungs: Evidence for an Antifibrotic Role in Bleomycin Injury." *The American Journal of Pathology* 180(5): 1963-1978.
221. Kauffmann, A., R. Gentleman and W. Huber (2009). "arrayQualityMetrics - A bioconductor package for quality assessment of microarray data." *Bioinformatics* 25(3): 415-416.
222. Kaul, B., J. S. Lee, N. Zhang, E. Vittinghoff, K. Sarmiento, H. R. Collard and M. A. Whooley (2021). "Epidemiology of Idiopathic Pulmonary Fibrosis among U.S. Veterans, 2010–2019." *Annals of the American Thoracic Society* 19(2): 196-203.
223. Khan, T., S. Dasgupta, N. Ghosh and K. Chaudhury (2021). "Proteomics in idiopathic pulmonary fibrosis: the quest for biomarkers." *Molecular Omics* 17(1): 43-58.
224. Kheir, F., A. Alkhatib, G. J. Berry, P. Daroca, L. Diethelm, R. Rampolla, S. Saito, D. L. Smith, D. Weill, M. Bateman, R. Abdelghani and J. A. Lasky (2020). "Using Bronchoscopic Lung Cryobiopsy and a Genomic Classifier in the Multidisciplinary Diagnosis of Diffuse Interstitial Lung Diseases." *CHEST* 158(5): 2015-2025.
225. Kim, D., J. M. Paggi, C. Park, C. Bennett and S. L. Salzberg (2019). "Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype." *Nature Biotechnology* 37(8): 907-915.
226. Kim, S., J. D. Herazo-Maya, D. D. Kang, B. M. Juan-Guardela, J. Tedrow, F. J. Martinez, F. C. Scirba, G. C. Tseng and N. Kaminski (2015). "Integrative phenotyping framework (iPF): integrative clustering of multiple omics data identifies novel lung disease subphenotypes." *BMC Genomics* 16(1): 924.
227. Kim, S. K., S. M. Jung, K.-S. Park and K.-J. Kim (2021). "Integrative analysis of lung molecular signatures reveals key drivers of idiopathic pulmonary fibrosis." *BMC Pulmonary Medicine* 21(1): 404.
228. Kim, S. Y., J. Diggans, D. Pankratz, J. Huang, M. Pagan, N. Sindy, E. Tom, J. Anderson, Y. Choi, D. A. Lynch, M. P. Steele, K. R. Flaherty, K. K. Brown, H. Farah, M. J. Bukstein, A. Pardo, M. Selman, P. J. Wolters, S. D. Nathan, T. V. Colby, J. L. Myers, A.-L. A. Katzenstein, G. Raghu and G. C. Kennedy (2015). "Classification of usual interstitial pneumonia in patients with interstitial lung disease: assessment of a machine learning approach using high-dimensional transcriptional data." *The Lancet Respiratory Medicine* 3(6): 473-482.
229. King, T. E., W. Z. Bradford, S. Castro-Bernardini, E. A. Fagan, I. Glaspole, M. K. Glassberg, E. Gorina, P. M. Hopkins, D. Kardatzke, L. Lancaster, D. J. Lederer, S. D. Nathan, C. A. Pereira, S. A. Sahn, R. Sussman, J. J. Swigris and P. W. Noble (2014). "A Phase 3 Trial of Pirfenidone in Patients with Idiopathic Pulmonary Fibrosis." *New England Journal of Medicine* 370(22): 2083-2092.
230. Kitanaka, N., R. Nakano, K. Sugiura, T. Kitanaka, S. Namba, T. Konno, T. Nakayama and H. Sugiya (2019). "Interleukin-1 β promotes interleukin-6 expression via ERK1/2 signaling pathway in canine dermal fibroblasts." *PLOS ONE* 14(7): e0220262.
231. Kjeldsen, L., A. H. Johnsen, H. Sengeløv and N. Borregaard (1993). "Isolation and primary structure of NGAL, a novel protein associated with human neutrophil gelatinase." *Journal of Biological Chemistry* 268(14): 10425-10432.
232. Klemm, C., C. Bruchhagen, A. van Krüchten, S. Niemann, B. Löffler, G. Peters, S. Ludwig and C. Ehrhardt (2017). "Mitogen-activated protein kinases (MAPKs) regulate IL-6 over-production during concomitant influenza virus and Staphylococcus aureus infection." *Scientific Reports* 7(1): 42473.
233. Knobloch, J., Y. Lin, J. Konradi, D. Jungck, J. Behr, J. Strauch, E. Stoelben and A. Koch (2013). "Inflammatory Responses of Airway Smooth Muscle Cells and Effects of Endothelin Receptor Antagonism." *American Journal of Respiratory Cell and Molecular Biology* 49(1): 114-127.
234. Kobayashi, Y., A. Tata, A. Konkimalla, H. Katsura, R. F. Lee, J. Ou, N. E. Banovich, J. A. Kropski and P. R. Tata (2020). "Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis." *Nature Cell Biology* 22(8): 934-946.

235. Kodama, Y., J. Mashima, E. Kaminuma, T. Gojobori, O. Ogasawara, T. Takagi, K. Okubo and Y. Nakamura (2012). "The DNA Data Bank of Japan launches a new resource, the DDBJ Omics Archive of functional genomics experiments." *Nucleic Acids Research* 40(D1): D38-D42.
236. Kodama, Y., M. Shumway, R. Leinonen and C. on behalf of the International Nucleotide Sequence Database (2012). "The sequence read archive: explosive growth of sequencing data." *Nucleic Acids Research* 40(D1): D54-D56.
237. Koga, Y., H. Tsurumaki, H. Aoki-Saito, M. Sato, M. Yatomi, K. Takehara and T. Hisada (2019) "Roles of Cyclic AMP Response Element Binding Activation in the ERK1/2 and p38 MAPK Signalling Pathway in Central Nervous System, Cardiovascular System, Osteoclast Differentiation and Mucin and Cytokine Production." *International Journal of Molecular Sciences* 20 DOI: 10.3390/ijms20061346.
238. Konishi, K., K. F. Gibson, K. O. Lindell, T. J. Richards, Y. Zhang, R. Dhir, M. Bisceglia, S. Gilbert, S. A. Yousem, J. W. Song, D. S. Kim and N. Kaminski (2009). "Gene Expression Profiles of Acute Exacerbations of Idiopathic Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 180(2): 167-175.
239. Kozomara, A., M. Birgaoanu and S. Griffiths-Jones (2019). "MiRBase: From microRNA sequences to function." *Nucleic Acids Research* 47(D1): D155-D162.
240. Krassowski, M., V. Das, S. K. Sahu and B. B. Misra (2020). "State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing." *Frontiers in Genetics* 11.
241. Kuppe, C., M. M. Ibrahim, J. Kranz, X. Zhang, S. Ziegler, J. Perales-Patón, J. Jansen, K. C. Reimer, J. R. Smith, R. Dobie, J. R. Wilson-Kanamori, M. Halder, Y. Xu, N. Kabgani, N. Kaesler, M. Klaus, L. Gernhold, V. G. Puelles, T. B. Huber, P. Boor, S. Menzel, R. M. Hoogenboezem, E. M. J. Bindels, J. Steffens, J. Floege, R. K. Schneider, J. Saez-Rodriguez, N. C. Henderson and R. Kramann (2021). "Decoding myofibroblast origins in human kidney fibrosis." *Nature* 589(7841): 281-286.
242. Kusko, R. L., J. F. Brothers, J. Tedrow, K. Pandit, L. Huleihel, C. Perdomo, G. Liu, B. Juan-Guardela, D. Kass, S. Zhang, M. Lenburg, F. Martinez, J. Quackenbush, F. Sciarba, A. Limper, M. Geraci, I. Yang, D. A. Schwartz, J. Beane, A. Spira and N. Kaminski (2016). "Integrated Genomics Reveals Convergent Transcriptomic Networks Underlying Chronic Obstructive Pulmonary Disease and Idiopathic Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 194(8): 948-960.
243. Lam, A. P., J. D. Herazo-Maya, J. A. Sennello, A. S. Flozak, S. Russell, G. M. Mutlu, G. R. S. Budinger, R. DasGupta, J. Varga, N. Kaminski and C. J. Gottardi (2014). "Wnt Coreceptor Lrp5 Is a Driver of Idiopathic Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 190(2): 185-195.
244. Lamb, J., E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander and T. R. Golub (2006). "The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease." *Science* 313(5795): 1929-1935.
245. Lambert, S. A., A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T. R. Hughes and M. T. Weirauch (2018). "The Human Transcription Factors." *Cell* 172(4): 650-665.
246. Langfelder, P. and S. Horvath (2008). "WGCNA: An R package for weighted correlation network analysis." *BMC Bioinformatics* 9.
247. Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." *Nature methods* 9(4): 357-359.
248. Larsson, O., D. Diebold, D. Fan, M. Peterson, R. S. Nho, P. B. Bitterman and C. A. Henke (2008). "Fibrotic Myofibroblasts Manifest Genome-Wide Derangements of Translational Control." *PLOS ONE* 3(9): e3220.
249. Law, C. W., Y. Chen, W. Shi and G. K. Smyth (2014). "voom: precision weights unlock linear model analysis tools for RNA-seq read counts." *Genome Biology* 15(2): R29.

250. Lawrence, M., W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan and V. J. Carey (2013). "Software for Computing and Annotating Genomic Ranges." *PLOS Computational Biology* 9(8): e1003118.
251. Lebel, M., D. O. Cliche, M. Charbonneau, D. Adam, E. Brochiero, C. M. Dubois and A. M. Cantin (2023) "Invadosome Formation by Lung Fibroblasts in Idiopathic Pulmonary Fibrosis." *International Journal of Molecular Sciences* 24 DOI: 10.3390/ijms24010499.
252. Lederer, D. J. and F. J. Martinez (2018). "Idiopathic Pulmonary Fibrosis." *New England Journal of Medicine* 378(19): 1811-1823.
253. Lee, J.-U., H. S. Cheong, E.-Y. Shim, D.-J. Bae, H. S. Chang, S.-T. Uh, Y. H. Kim, J.-S. Park, B. Lee, H. D. Shin and C.-S. Park (2017). "Gene profile of fibroblasts identify relation of CCL8 with idiopathic pulmonary fibrosis." *Respiratory Research* 18(1): 3.
254. Lee, J.-U., J.-H. Son, E.-Y. Shim, H. S. Cheong, S.-W. Shin, H. D. Shin, A. R. Baek, S. Ryu, C.-S. Park, H. S. Chang and J.-S. Park (2019). "Global DNA Methylation Pattern of Fibroblasts in Idiopathic Pulmonary Fibrosis." *DNA and Cell Biology* 38(9): 905-914.
255. Lee, J., I. Arisi, E. Puxeddu, L. K. Mramba, M. Amicosante, C. M. Swaisgood, M. Pallante, M. L. Brantly, C. M. Sköld and C. Saltini (2018). "Bronchoalveolar lavage (BAL) cells in idiopathic pulmonary fibrosis express a complex pro-inflammatory, pro-repair, angiogenic activation pattern, likely associated with macrophage iron accumulation." *PLOS ONE* 13(4): e0194803.
256. Lee, P. J., X. Zhang, P. Shan, B. Ma, C. G. Lee, R. J. Homer, Z. Zhu, M. Rincon, B. T. Mossman and J. A. Elias (2006). "ERK1/2 mitogen-activated protein kinase selectively mediates IL-13-induced lung inflammation and remodeling in vivo." *The Journal of Clinical Investigation* 116(1): 163-173.
257. Lemée, J.-M., A. Clavreul, M. Aubry, E. Com, M. de Tayrac, J. Mosser and P. Menei (2018). "Integration of transcriptome and proteome profiles in glioblastoma: looking for the missing link." *BMC Molecular Biology* 19(1): 13.
258. Levy, S. E. and R. M. Myers (2016). "Advancements in Next-Generation Sequencing." *Annual Review of Genomics and Human Genetics* 17(1): 95-115.
259. Ley, R. E., F. Bäckhed, P. Turnbaugh, C. A. Lozupone, R. D. Knight and J. I. Gordon (2005). "Obesity alters gut microbial ecology." *Proceedings of the National Academy of Sciences* 102(31): 11070-11075.
260. Ley, R. E., P. J. Turnbaugh, S. Klein and J. I. Gordon (2006). "Human gut microbes associated with obesity." *Nature* 444(7122): 1022-1023.
261. Li, B., S. M. Clohisey, B. S. Chia, B. Wang, A. Cui, T. Eisenhaure, L. D. Schweitzer, P. Hoover, N. J. Parkinson, A. Nachshon, N. Smith, T. Regan, D. Farr, M. U. Gutmann, S. I. Bukhari, A. Law, M. Sangesland, I. Gat-Viks, P. Digard, S. Vasudevan, D. Lingwood, D. H. Dockrell, J. G. Doench, J. K. Baillie and N. Hacohen (2020). "Genome-wide CRISPR screen identifies host dependency factors for influenza A virus infection." *Nature Communications* 11(1): 164.
262. Li, C. M.-C., G. Chen, T. L. Dayton, C. Kim-Kiselak, S. Hoersch, C. A. Whittaker, R. T. Bronson, D. G. Beer, M. M. Winslow and T. Jacks (2013). "Differential Tks5 isoform expression contributes to metastatic invasion of lung adenocarcinoma." *Genes & Development* 27(14): 1557-1567.
263. Li, D., Y. Liu and B. Wang (2020). "Identification of transcriptomic markers for developing idiopathic pulmonary fibrosis : an integrative analysis of gene expression profiles." *International Journal of Clinical and Experimental Pathology* 13(7): 1698-1706.
264. Li, H., C. Zhao, Y. Tian, J. Lu, G. Zhang, S. Liang, D. Chen, X. Liu, W. Kuang and M. Zhu (2020). "Src family kinases and pulmonary fibrosis: A review." *Biomedicine & Pharmacotherapy* 127: 110183.
265. Li, R., Y. Wang, X. Song, W. Sun, J. Zhang, Y. Liu, H. Li, C. Meng, J. Zhang, Q. Zheng and C. Lv (2018). "Potential regulatory role of circular RNA in idiopathic pulmonary fibrosis." *Int J Mol Med* 42(6): 3256-3268.

266. Li, S., J. Geng, X. Xu, X. Huang, D. Leng, D. Jiang, J. Liang, C. Wang, D. Jiang and H. Dai (2016). "miR-130b-3p Modulates Epithelial-Mesenchymal Crosstalk in Lung Fibrosis by Targeting IGF-1." *PLOS ONE* 11(3): e0150418.
267. Li, X., X. Sun, C. Kan, B. Chen, N. Qu, N. Hou, Y. Liu and F. Han (2022). "COL1A1: A novel oncogenic gene and therapeutic target in malignancies." *Pathology - Research and Practice* 236: 154013.
268. Li, Y., D. Jiang, J. Liang, E. B. Meltzer, A. Gray, R. Miura, L. Wogensen, Y. Yamaguchi and P. W. Noble (2011). "Severe lung fibrosis requires an invasive fibroblast phenotype regulated by hyaluronan and CD44." *Journal of Experimental Medicine* 208(7): 1459-1471.
269. Li, Z., S. Wang, H. Zhao, P. Yan, H. Yuan, M. Zhao, R. Wan, G. Yu and L. Wang (2023). "Artificial neural network identified the significant genes to distinguish Idiopathic pulmonary fibrosis." *Scientific Reports* 13(1): 1225.
270. Liberzon, A., A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo and J. P. Mesirov (2011). "Molecular signatures database (MSigDB) 3.0." *Bioinformatics* 27(12): 1739-1740.
271. Lindahl, G. E., C. J. W. Stock, X. Shi-Wen, P. Leoni, P. Sestini, S. L. Howat, G. Bou-Gharios, A. G. Nicholson, C. P. Denton, J. C. Grutters, T. M. Maher, A. U. Wells, D. J. Abraham and E. A. Renzoni (2013). "Microarray profiling reveals suppressed interferon stimulated gene program in fibroblasts from scleroderma-associated interstitial lung disease." *Respiratory Research* 14(1): 80.
272. Litvin, M., J. C. Yoon, J. Leey Casella, S. M. Blackman and A. L. Brennan (2019). "Energy balance and obesity in individuals with cystic fibrosis." *Journal of Cystic Fibrosis* 18: S38-S47.
273. Liu, T., L. Zhang, D. Joo and S.-C. Sun (2017). "NF- κ B signaling in inflammation." *Signal Transduction and Targeted Therapy* 2(1): 17023.
274. Liu, X., Y. Peng and J. Wang (2020). "Integrative analysis of DNA methylation and gene expression profiles identified potential breast cancer-specific diagnostic markers." *Bioscience Reports* 40(5): BSR20201053.
275. Liu, X., S. C. Rowan, J. Liang, C. Yao, G. Huang, N. Deng, T. Xie, D. Wu, Y. Wang, A. Burman, T. Parimon, Z. Borok, P. Chen, W. C. Parks, C. M. Hogaboam, S. S. Weigt, J. Belperio, B. R. Stripp, P. W. Noble and D. Jiang (2021). "Categorization of lung mesenchymal cells in development and fibrosis." *iScience* 24(6).
276. Lloyd, C. M., A. W. Minto, M. E. Dorf, A. Proudfoot, T. N. C. Wells, D. J. Salant and J.-C. Gutierrez-Ramos (1997). "RANTES and Monocyte Chemoattractant Protein-1 (MCP-1) Play an Important Role in the Inflammatory Phase of Crescentic Nephritis, but Only MCP-1 Is Involved in Crescent Formation and Interstitial Fibrosis." *Journal of Experimental Medicine* 185(7): 1371-1380.
277. Lock, P., C. L. Abram, T. Gibson and S. A. Courtneidge (1998). "A new method for isolating tyrosine kinase substrates used to identify Fish, an SH3 and PX domain-containing protein, and Src substrate." *The EMBO Journal* 17(15): 4346-4357.
278. Lockstone, H. E., S. Sanderson, N. Kulakova, D. Baban, A. Leonard, W. L. Kok, S. McGowan, A. J. McMichael and L.-P. Ho (2010). "Gene Set Analysis of Lung Samples Provides Insight into Pathogenesis of Progressive, Fibrotic Pulmonary Sarcoidosis." *American Journal of Respiratory and Critical Care Medicine* 181(12): 1367-1375.
279. Loganathan, T. and G. P. Doss C (2023). "Non-coding RNAs in human health and disease: potential function as biomarkers and therapeutic targets." *Functional & Integrative Genomics* 23(1): 33.
280. Lokk, K., V. Modhukur, B. Rajashekar, K. Märtens, R. Mägi, R. Kolde, M. Koltšina, T. K. Nilsson, J. Vilo, A. Salumets and N. Tõnisson (2014). "DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns." *Genome Biology* 15(4): 3248.
281. Loretelli, C., F. Rocchio, F. D'Addio, M. Ben Nasr, E. Castillo-Leon, S. Dellepiane, A. Vergani, A. Abdelsalam, E. Assi, A. Maestroni, V. Uselli, R. Bassi, I. Pastore, J. Yang, B. El Essawy, K. M. Elased, G. P. Fadini, E. Ippolito, A. J. Seelam, M. Pezzolesi, D. Corradi, G. V. Zuccotti, M. Gallieni,

- M. Allegretti, M. A. Niewczasz and P. Fiorina (2021). "The IL-8-CXCR1/2 axis contributes to diabetic kidney disease." *Metabolism - Clinical and Experimental* 121.
282. Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology* 15(12): 1-21.
283. Lovgren, A. K., J. J. Kovacs, T. Xie, E. N. Potts, Y. Li, W. M. Foster, J. Liang, E. B. Meltzer, D. Jiang, R. J. Lefkowitz and P. W. Noble (2011). " β -Arrestin Deficiency Protects Against Pulmonary Fibrosis in Mice and Prevents Fibroblast Invasion of Extracellular Matrix." *Science Translational Medicine* 3(74): 74ra23-74ra23.
284. Łukasiewicz, S., M. Czeczulewski, A. Forma, J. Baj, R. Sitarz and A. Stanisławek (2021). "Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review." *Cancers* 13(17).
285. Lundberg, S. M., G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal and S.-I. Lee (2020). "From local explanations to global understanding with explainable AI for trees." *Nature Machine Intelligence* 2(1): 56-67.
286. Lundberg, S. M. and S.-I. Lee (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, California, USA, Curran Associates Inc.: 4768–4777.
287. Luzina, I. G., M. V. Salcedo, M. L. Rojas-Peña, A. E. Wyman, J. R. Galvin, A. Sachdeva, A. Clerman, J. Kim, T. J. Franks, E. J. Britt, J. D. Hasday, S. M. Pham, A. P. Burke, N. W. Todd and S. P. Atamas (2018). "Transcriptomic evidence of immune activation in macroscopically normal-appearing and scarred lung tissues in idiopathic pulmonary fibrosis." *Cellular Immunology* 325: 1-13.
288. Lv, W., G. W. Booz, Y. Wang, F. Fan and R. J. Roman (2018). "Inflammation and renal fibrosis: Recent developments on key signaling molecules as potential therapeutic targets." *European Journal of Pharmacology* 820: 65-76.
289. Lynch, D. A., N. Sverzellati, W. D. Travis, K. K. Brown, T. V. Colby, J. R. Galvin, J. G. Goldin, D. M. Hansell, Y. Inoue, T. Johkoh, A. G. Nicholson, S. L. Knight, S. Raof, L. Richeldi, C. J. Ryerson, J. H. Ryu and A. U. Wells (2018). "Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper." *The Lancet Respiratory Medicine* 6(2): 138-153.
290. Lynch, J. P., III, R. Sagar, S. S. Weigt, D. A. Zisman and E. S. White (2006). "Usual Interstitial Pneumonia." *Semin Respir Crit Care Med* 27(06): 634-651.
291. Mackintosh, J. A., S. R. Desai, H. Adamali, K. Patel, F. Chua, A. Devaraj, V. Kouranos, M. Kokosi, G. Margaritopoulos, E. A. Renzoni, A. U. Wells, P. L. Molyneaux, S. Kumar, T. M. Maher and P. M. George (2019). "In patients with idiopathic pulmonary fibrosis the presence of hiatus hernia is associated with disease progression and mortality." *European Respiratory Journal* 53(5): 1802412.
292. Maghsoudloo, M., S. Azimzadeh Jamalkandi, A. Najafi and A. Masoudi-Nejad (2020). "An efficient hybrid feature selection method to identify potential biomarkers in common chronic lung inflammatory diseases." *Genomics* 112(5): 3284-3293.
293. Magkrioti, C. and V. Aidinis (2013). "Autotaxin and lysophosphatidic acid signalling in lung pathophysiology." *World Journal of Respirology* 3(3): 77-103.
294. Magkrioti, C., G. Antonopoulou, D. Fanidis, V. Pliaka, T. Sakellaropoulos, L. G. Alexopoulos, C. Ullmer and V. Aidinis (2022) "Lysophosphatidic Acid Is a Proinflammatory Stimulus of Renal Tubular Epithelial Cells." *International Journal of Molecular Sciences* 23 DOI: 10.3390/ijms23137452.
295. Magkrioti, C., A. Galaris, P. Kanellopoulou, E.-A. Stylianaki, E. Kaffe and V. Aidinis (2019). "Autotaxin and chronic inflammatory diseases." *Journal of Autoimmunity* 104: 102327.
296. Magkrioti, C., N. Oikonomou, E. Kaffe, M.-A. Mouratis, N. Xylourgidis, I. Barbayianni, P. Megadoukas, V. Harokopos, C. Valavanis, J. Chun, A. Kosma, G. T. Stathopoulos, E. Bouros, D.

- Bouros, K. Syrigos and V. Aidinis (2018). "The Autotaxin—Lysophosphatidic Acid Axis Promotes Lung Carcinogenesis." *Cancer Research* 78(13): 3634-3644.
297. Maher, T. M., I. C. Evans, S. E. Bottoms, P. F. Mercer, A. J. Thorley, A. G. Nicholson, G. J. Laurent, T. D. Tetley, R. C. Chambers and R. J. McAnulty (2010). "Diminished Prostaglandin E2 Contributes to the Apoptosis Paradox in Idiopathic Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 182(1): 73-82.
298. Maier, T., M. Güell and L. Serrano (2009). "Correlation of mRNA and protein in complex biological samples." *FEBS Letters* 583(24): 3966-3973.
299. Manitsopoulos, N., V. Aidinis, K. Perreas, S. Orfanos and A. Kotanidou (2017). "The effects of tpl2 inhibition in Ventilator-induced lung injury." *European Respiratory Journal* 50(suppl 61): PA349.
300. Manzoni, C., D. A. Kia, J. Vandrovцова, J. Hardy, N. W. Wood, P. A. Lewis and R. Ferrari (2018). "Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences." *Briefings in Bioinformatics* 19(2): 286-302.
301. Mårtensson, J., S. Xu, M. Bell, C.-R. Martling and P. Venge (2012). "Immunoassays distinguishing between HNL/NGAL released in urine from kidney epithelial cells and neutrophils." *Clinica Chimica Acta* 413(19): 1661-1667.
302. Mayer, T. Z., F. A. Simard, A. Cloutier, H. Vardhan, C. M. Dubois and P. P. McDonald (2013). "The p38-MSK1 Signaling Cascade Influences Cytokine Production through CREB and C/EBP Factors in Human Neutrophils." *The Journal of Immunology* 191(8): 4299-4307.
303. Mayr, C. H., L. M. Simon, G. Leuschner, M. Ansari, J. Schniering, P. E. Geyer, I. Angelidis, M. Strunz, P. Singh, N. Kneidinger, F. Reichenberger, E. Silbernagel, S. Böhm, H. Adler, M. Lindner, B. Maurer, A. Hilgendorff, A. Prasse, J. Behr, M. Mann, O. Eickelberg, F. J. Theis and H. B. Schiller (2021). "Integrative analysis of cell state changes in lung fibrosis with peripheral protein biomarkers." *EMBO Molecular Medicine* 13(4): e12871.
304. McCarthy, D. J., Y. Chen and G. K. Smyth (2012). "Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation." *Nucleic Acids Research* 40(10): 4288-4297.
305. McDonough, J. E., N. Kaminski, B. Thienpont, J. C. Hogg, B. M. Vanaudenaerde and W. A. Wuyts (2019). "Gene correlation network analysis to identify regulatory factors in idiopathic pulmonary fibrosis." *Thorax* 74(2): 132.
306. McInnes, L., H. John and M. James (2020). "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." *arXiv*.
307. Mehaffey, E. and D. S. A. Majid (2017). "Tumor necrosis factor- α , kidney function, and hypertension." *American Journal of Physiology-Renal Physiology* 313(4): F1005-F1008.
308. Melboucy-Belkhir, S., P. Pradère, S. Tadbiri, S. Habib, A. Bacrot, S. Brayer, B. Mari, V. Besnard, A. Mailleux, A. Guenther, Y. Castier, H. Mal, B. Crestani and L. Plantier (2014). "Forkhead Box F1 represses cell growth and inhibits COL1 and ARPC2 expression in lung fibroblasts in vitro." *American Journal of Physiology-Lung Cellular and Molecular Physiology* 307(11): L838-L847.
309. Melotti, P., E. Nicolis, A. Tamanini, R. Rolfini, A. Pavirani and G. Cabrini (2001). "Activation of NF- κ B mediates ICAM-1 induction in respiratory cells exposed to an adenovirus-derived vector." *Gene Therapy* 8(18): 1436-1442.
310. Memet, I., E. Tsalkidou, A. K. Tsaroucha, M. Lambropoulou, E. Chatzaki, G. Trypsianis, D. Schizas, M. Pitiakoudis and C. Simopoulos (2018). "Autotaxin Expression in Hepatocellular Carcinoma." *Journal of Investigative Surgery* 31(5): 359-365.
311. Meng, G., X. Tang, Z. Yang, Y. Y. Zhao, J. M. Curtis, T. P. W. McMullen and D. N. Brindley (2019). "Dexamethasone decreases the autotaxin-lysophosphatidate-inflammatory axis in adipose tissue: implications for the metabolic syndrome and breast cancer." *The FASEB Journal* 33(2): 1899-1910.

312. Michopoulou, A., M. Montmasson, C. Garnier, E. Lambert, G. Dayan and P. Rousselle (2020). "A novel mechanism in wound healing: Laminin 332 drives MMP9/14 activity by recruiting syndecan-1 and CD44." *Matrix Biology* 94: 1-17.
313. Milosevic, J., K. Pandit, M. Magister, E. Rabinovich, D. C. Ellwanger, G. Yu, L. J. Vuga, B. Weksler, P. V. Benos, K. F. Gibson, M. McMillan, M. Kahn and N. Kaminski (2012). "Profibrotic Role of miR-154 in Pulmonary Fibrosis." *American Journal of Respiratory Cell and Molecular Biology* 47(6): 879-887.
314. Mirzoyan, K., A. Baïotto, A. Dupuy, D. Marsal, C. Denis, C. Vinel, P. Sicard, J. Bertrand-Michel, J.-L. Bascands, J. P. Schanstra, J. Klein and J.-S. Saulnier-Blache (2016). "Increased urinary lysophosphatidic acid in mouse with subtotal nephrectomy: potential involvement in chronic kidney disease." *Journal of Physiology and Biochemistry* 72(4): 803-812.
315. Misharin, A. V., L. Morales-Nebreda, P. A. Reyfman, C. M. Cuda, J. M. Walter, A. C. McQuattie-Pimentel, C.-I. Chen, K. R. Anekalla, N. Joshi, K. J. N. Williams, H. Abdala-Valencia, T. J. Yacoub, M. Chi, S. Chiu, F. J. Gonzalez-Gonzalez, K. Gates, A. P. Lam, T. T. Nicholson, P. J. Homan, S. Soberanes, S. Dominguez, V. K. Morgan, R. Saber, A. Shaffer, M. Hinchcliff, S. A. Marshall, A. Bharat, S. Berdnikovs, S. M. Bhorade, E. T. Bartom, R. I. Morimoto, W. E. Balch, J. I. Sznajder, N. S. Chandel, G. M. Mutlu, M. Jain, C. J. Gottardi, B. D. Singer, K. M. Ridge, N. Bagheri, A. Shilatifard, G. R. S. Budinger and H. Perlman (2017). "Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span." *Journal of Experimental Medicine* 214(8): 2387-2404.
316. Moll, P., M. Ante, A. Seitz and T. Reda (2014). "QuantSeq 3' mRNA sequencing for RNA quantification." *Nature Methods* 11(12): i-iii.
317. Molyneaux, P. L., M. J. Cox, S. A. G. Willis-Owen, P. Mallia, K. E. Russell, A.-M. Russell, E. Murphy, S. L. Johnston, D. A. Schwartz, A. U. Wells, W. O. C. Cookson, T. M. Maher and M. F. Moffatt (2014). "The Role of Bacteria in the Pathogenesis and Progression of Idiopathic Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 190(8): 906-913.
318. Molyneaux, P. L., S. A. G. Willis-Owen, M. J. Cox, P. James, S. Cowman, M. Loebinger, A. Blanchard, L. M. Edwards, C. Stock, C. Daccord, E. A. Renzoni, A. U. Wells, M. F. Moffatt, W. O. C. Cookson and T. M. Maher (2017). "Host-Microbial Interactions in Idiopathic Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 195(12): 1640-1650.
319. Morais, L. H., H. L. Schreiber and S. K. Mazmanian (2021). "The gut microbiota-brain axis in behaviour and brain disorders." *Nature Reviews Microbiology* 19(4): 241-255.
320. Morse, C., T. Tabib, J. Sembrat, K. L. Buschur, H. T. Bittar, E. Valenzi, Y. Jiang, D. J. Kass, K. Gibson, W. Chen, A. Mora, P. V. Benos, M. Rojas and R. Lafyatis (2019). "Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis." *European Respiratory Journal* 54(2): 1802441.
321. Moulos, P. and P. Hatzis (2015). "Systematic integration of RNA-Seq statistical algorithms for accurate detection of differential gene expression patterns." *Nucleic Acids Research* 43(4): e25-e25.
322. Mouratis, M. A. and V. Aidinis (2011). "Modeling pulmonary fibrosis with bleomycin." *Current Opinion in Pulmonary Medicine* 17(5).
323. Mukherjee, S., W. Sheng, A. Michkov, K. Sriarm, R. Sun, A. Dvorkin-Gheva, P. A. Insel and L. J. Janssen (2019). "Prostaglandin E2 inhibits profibrotic function of human pulmonary fibroblasts by disrupting Ca²⁺ signaling." *American Journal of Physiology-Lung Cellular and Molecular Physiology* 316(5): L810-L821.
324. Mura, M., M. Anraku, Z. Yun, K. McRae, M. Liu, T. K. Waddell, L. G. Singer, J. T. Granton, S. Keshavjee and M. de Perrot (2012). "Gene Expression Profiling in the Lungs of Patients With Pulmonary Hypertension Associated With Pulmonary Fibrosis." *Chest* 141(3): 661-673.
325. Murphy, D. A. and S. A. Courtneidge (2011). "The 'ins' and 'outs' of podosomes and invadopodia: characteristics, formation and function." *Nature Reviews Molecular Cell Biology* 12(7): 413-426.

326. Nakahara, Y., N. Hashimoto, K. Sakamoto, A. Enomoto, T. S. Adams, T. Yokoi, N. Omote, S. Poli, A. Ando, K. Wakahara, A. Suzuki, M. Inoue, A. Hara, Y. Mizutani, K. Imaizumi, T. Kawabe, I. O. Rosas, M. Takahashi, N. Kaminski and Y. Hasegawa (2021). "Fibroblasts positive for meflin have anti-fibrotic property in pulmonary fibrosis." *European Respiratory Journal*: 2003397.
327. Namba, S., R. Nakano, T. Kitanaka, N. Kitanaka, T. Nakayama and H. Sugiya (2017). "ERK2 and JNK1 contribute to TNF- α -induced IL-8 expression in synovial fibroblasts." *PLOS ONE* 12(8): e0182923.
328. Nance, T., K. S. Smith, V. Anaya, R. Richardson, L. Ho, M. Pala, S. Mostafavi, A. Battle, C. Feghali-Bostwick, G. Rosen and S. B. Montgomery (2014). "Transcriptome Analysis Reveals Differential Splicing Events in IPF Lung Tissue." *PLOS ONE* 9(3): e92111.
329. Nathan, S., H. Zhang, M. Andreoli, P. L. Leopold and R. G. Crystal (2021). "CREB-dependent LPA-induced signaling initiates a pro-fibrotic feedback loop between small airway basal cells and fibroblasts." *Respiratory Research* 22(1): 97.
330. Nathan, S. D., O. A. Shlobin, N. Weir, S. Ahmad, J. M. Kaldjob, E. Battle, M. J. Sheridan and R. M. du Bois (2011). "Long-term Course and Prognosis of Idiopathic Pulmonary Fibrosis in the New Millennium." *Chest* 140(1): 221-229.
331. Negreros, M., J. S. Hagood, C. R. Espinoza, Y. I. Balderas-Martínez, M. Selman and A. Pardo (2019). "Transforming growth factor beta 1 induces methylation changes in lung fibroblasts." *PLOS ONE* 14(10): e0223512.
332. Nema, R., A. Shrivastava and A. Kumar (2021). "Prognostic role of lipid phosphate phosphatases in non-smoker, lung adenocarcinoma patients." *Computers in Biology and Medicine* 129: 104141.
333. Newton, C. A., D. Zhang, J. M. Oldham, J. Kozlitina, S.-F. Ma, F. J. Martinez, G. Raghu, I. Noth and C. K. Garcia (2018). "Telomere Length and Use of Immunosuppressive Medications in Idiopathic Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 200(3): 336-347.
334. Nickolas, T. L., C. S. Forster, M. E. Sise, N. Barasch, D. S.-D. Valle, M. Viltard, C. Buchen, S. Kupferman, M. L. Carnevali, M. Bennett, S. Mattei, A. Bovino, L. Argentiero, A. Magnano, P. Devarajan, K. Mori, H. Erdjument-Bromage, P. Tempst, L. Allegri and J. Barasch (2012). "NGAL (Lcn2) monomer is associated with tubulointerstitial damage in chronic kidney disease." *Kidney International* 82(6): 718-722.
335. Nikitopoulou, I., D. Fanidis, K. Ntatsoulis, P. Moulos, G. Mpekoulis, M. Evangelidou, A. G. Vassiliou, V. Dimakopoulou, E. Jahaj, S. Tshipilis, S. E. Orfanos, I. Dimopoulou, E. Angelakis, K. Akinosoglou, N. Vassilaki, A. Tzouveleakis, A. Kotanidou and V. Aidinis (2021). "Increased Autotaxin Levels in Severe COVID-19, Correlating with IL-6 Levels, Endothelial Dysfunction Biomarkers, and Impaired Functions of Dendritic Cells." *International Journal of Molecular Sciences* 22(18).
336. Nikitopoulou, I., N. Oikonomou, E. Karouzakis, I. Sevastou, N. Nikolaidou-Katsaridou, Z. Zhao, V. Mersinias, M. Armaka, Y. Xu, M. Masu, G. B. Mills, S. Gay, G. Kollias and V. Aidinis (2012). "Autotaxin expression from synovial fibroblasts is essential for the pathogenesis of modeled arthritis." *Journal of Experimental Medicine* 209(5): 925-933.
337. Ninou, I., E. Kaffe, S. Müller, D. C. Budd, C. S. Stevenson, C. Ullmer and V. Aidinis (2018). "Pharmacologic targeting of the ATX/LPA axis attenuates bleomycin-induced pulmonary fibrosis." *Pulmonary Pharmacology & Therapeutics* 52: 32-40.
338. Ninou, I., C. Magkrioti and V. Aidinis (2018). "Autotaxin in Pathophysiology and Pulmonary Fibrosis." *Frontiers in Medicine* 5: 180.
339. Noguchi, S., A. Saito, Y. Mikami, H. Urushiyama, M. Horie, H. Matsuzaki, H. Takeshima, K. Makita, N. Miyashita, A. Mitani, T. Jo, Y. Yamauchi, Y. Terasaki and T. Nagase (2017). "TAZ contributes to pulmonary fibrosis by activating profibrotic functions of lung fibroblasts." *Scientific Reports* 7(1): 42595.

340. Norman, K. C., B. B. Moore, K. B. Arnold and D. N. O'Dwyer (2018). "Proteomics: Clinical and research applications in respiratory diseases." *Respirology* 23(11): 993-1003.
341. Noth, I., Y. Zhang, S.-F. Ma, C. Flores, M. Barber, Y. Huang, S. M. Broderick, M. S. Wade, P. Hysi, J. Scurba, T. J. Richards, B. M. Juan-Guardela, R. Vij, M. K. Han, F. J. Martinez, K. Kossen, S. D. Seiwert, J. D. Christie, D. Nicolae, N. Kaminski and J. G. N. Garcia (2013). "Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study." *The Lancet Respiratory Medicine* 1(4): 309-317.
342. O'Leary, N. A., M. W. Wright, J. R. Brister, S. Ciufu, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy and K. D. Pruitt (2016). "Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation." *Nucleic Acids Research* 44(D1): D733-D745.
343. Odani, H., J. Asami, A. Ishii, K. Oide, T. Sudo, A. Nakamura, N. Miyata, N. Otsuka, K. Maeda and J. Nakagawa (2008). "Suppression of Renal α -Dicarbonyl Compounds Generated following Ureteral Obstruction by Kidney-Specific α -Dicarbonyl/l-Xylulose Reductase." *Annals of the New York Academy of Sciences* 1126(1): 320-324.
344. Oh, D.-J., H. R. Kim, M.-K. Lee and Y. S. Woo (2013). "Profile of Human β -Defensins 1,2 and Proinflammatory Cytokines (TNF- α , IL-6) in Patients with Chronic Kidney Disease." *Kidney and Blood Pressure Research* 37(6): 602-610.
345. Oikonomou, N., M.-A. Mouratis, A. Tzouvelekis, E. Kaffe, C. Valavanis, G. Vilaras, A. Karameris, G. D. Prestwich, D. Bouros and V. Aidinis (2012). "Pulmonary Autotaxin Expression Contributes to the Pathogenesis of Pulmonary Fibrosis." *American Journal of Respiratory Cell and Molecular Biology* 47(5): 566-574.
346. Olsen, T. K. and N. Baryawno (2018). "Introduction to Single-Cell RNA Sequencing." *Current Protocols in Molecular Biology* 122(1): e57.
347. Orejudo, M., R. R. Rodrigues-Diez, R. Rodrigues-Diez, A. Garcia-Redondo, L. Santos-Sánchez, J. Rández-Garbayo, P. Cannata-Ortiz, A. M. Ramos, A. Ortiz, R. Selgas, S. Mezzano, C. Lavoiz and M. Ruiz-Ortega (2019). "Interleukin 17A Participates in Renal Inflammation Associated to Experimental and Human Hypertension." *Frontiers in Pharmacology* 10.
348. Ortiz, A., A. B. Sanz, B. M. García, J. A. Moreno, M. D. S. Niño, J. L. Martín-Ventura, J. Egido and L. M. Blanco-Colio (2009). "Considering TWEAK as a target for therapy in renal and vascular injury." *Cytokine & Growth Factor Reviews* 20(3): 251-258.
349. Osuchowski, M. F., M. S. Winkler, T. Skirecki, S. Cajander, M. Shankar-Hari, G. Lachmann, G. Monneret, F. Venet, M. Bauer, F. M. Brunkhorst, S. Weis, A. Garcia-Salido, M. Kox, J.-M. Cavillon, F. Uhle, M. A. Weigand, S. B. Flohé, W. J. Wiersinga, R. Almansa, A. de la Fuente, I. Martin-Loeches, C. Meisel, T. Spinetti, J. C. Schefold, C. Cilloniz, A. Torres, E. J. Giamarellos-Bourboulis, R. Ferrer, M. Girardis, A. Cossarizza, M. G. Netea, T. van der Poll, J. F. Bermejo-Martín and I. Rubio (2021). "The COVID-19 puzzle: deciphering pathophysiology and phenotypes of a new disease entity." *The Lancet Respiratory Medicine* 9(6): 622-642.
350. Ota, M., Y. Nagafuchi, H. Hatano, K. Ishigaki, C. Terao, Y. Takeshima, H. Yanaoka, S. Kobayashi, M. Okubo, H. Shirai, Y. Sugimori, J. Maeda, M. Nakano, S. Yamada, R. Yoshida, H. Tsuchiya, Y. Tsuchida, S. Akizuki, H. Yoshifuji, K. Ohmura, T. Mimori, K. Yoshida, D. Kurosaka, M. Okada, K. Setoguchi, H. Kaneko, N. Ban, N. Yabuki, K. Matsuki, H. Mutoh, S. Oyama, M. Okazaki, H. Tsunoda, Y. Iwasaki, S. Sumitomo, H. Shoda, Y. Kochi, Y. Okada, K. Yamamoto, T. Okamura and K. Fujio (2021). "Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases." *Cell* 184(11): 3006-3021.e3017.

351. Pan, J., J. Hofmanninger, K.-H. Nanning, F. Prayer, S. Röhrich, N. Sverzellati, V. Poletti, S. Tomassetti, M. Weber, H. Prosch and G. Langa (2023). "Unsupervised machine learning identifies predictive progression markers of IPF." *European Radiology* 33(2): 925-935.
352. Panagopoulou, M., A. Cheretaki, M. Karaglani, I. Balgkouranidou, E. Biziota, K. Amarantidis, N. Xenidis, S. Kakolyris, S. Baritaki and E. Chatzaki (2021) "Methylation Status of Corticotropin-Releasing Factor (CRF) Receptor Genes in Colorectal Cancer." *Journal of Clinical Medicine* 10 DOI: 10.3390/jcm10122680.
353. Panagopoulou, M., A. Drosouni, D. Fanidis, M. Karaglani, I. Balgkouranidou, N. Xenidis, V. Aidinis and E. Chatzaki (2022). "ENPP2 Promoter Methylation Correlates with Decreased Gene Expression in Breast Cancer: Implementation as a Liquid Biopsy Biomarker." *International Journal of Molecular Sciences* 23(7).
354. Panagopoulou, M., D. Fanidis, V. Aidinis and E. Chatzaki (2021). "ENPP2 Methylation in Health and Cancer." *International Journal of Molecular Sciences* 22(21).
355. Panagopoulou, M., M. Karaglani, I. Balgkouranidou, E. Biziota, T. Koukaki, E. Karamitrousis, E. Nena, I. Tsamardinos, G. Kolios, E. Lianidou, S. Kakolyris and E. Chatzaki (2019). "Circulating cell-free DNA in breast cancer: size profiling, levels, and methylation patterns lead to prognostic and predictive classifiers." *Oncogene* 38(18): 3387-3401.
356. Panagopoulou, M., M. Karaglani, I. Balgkouranidou, C. Pantazi, G. Kolios, S. Kakolyris and E. Chatzaki (2019). "Circulating cell-free DNA release in vitro: kinetics, size profiling, and cancer-related gene methylation." *Journal of Cellular Physiology* 234(8): 14079-14089.
357. Pandit, K. V., D. Corcoran, H. Yousef, M. Yarlagadda, A. Tzouvelekis, K. F. Gibson, K. Konishi, S. A. Yousem, M. Singh, D. Handley, T. Richards, M. Selman, S. C. Watkins, A. Pardo, A. Ben-Yehudah, D. Bouros, O. Eickelberg, P. Ray, P. V. Benos and N. Kaminski (2010). "Inhibition and Role of let-7d in Idiopathic Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 182(2): 220-229.
358. Pankratz, D. G., Y. Choi, U. Imtiaz, G. M. Fedorowicz, J. D. Anderson, T. V. Colby, J. L. Myers, D. A. Lynch, K. K. Brown, K. R. Flaherty, M. P. Steele, S. D. Groshong, G. Raghu, N. M. Barth, P. S. Walsh, J. Huang, G. C. Kennedy and F. J. Martinez (2017). "Usual Interstitial Pneumonia Can Be Detected in Transbronchial Biopsies Using Machine Learning." *Annals of the American Thoracic Society* 14(11): 1646-1654.
359. Pardo, A., K. Gibson, J. Cisneros, T. J. Richards, Y. Yang, C. Becerril, S. Yousem, I. Herrera, V. Ruiz, M. Selman and N. Kaminski (2005). "Up-Regulation and Profibrotic Role of Osteopontin in Human Idiopathic Pulmonary Fibrosis." *PLOS Medicine* 2(9): e251.
360. Pardo, A. and M. Selman (2016). "Lung Fibroblasts, Aging, and Idiopathic Pulmonary Fibrosis." *Annals of the American Thoracic Society* 13(Supplement_5): S417-S421.
361. Parker, M. W., D. Rossi, M. Peterson, K. Smith, K. Sikström, E. S. White, J. E. Connett, C. A. Henke, O. Larsson and P. B. Bitterman (2014). "Fibrotic extracellular matrix activates a profibrotic positive feedback loop." *The Journal of Clinical Investigation* 124(4): 1622-1635.
362. Parris, T. Z., A. Kovács, S. Hajizadeh, S. Nemes, M. Semaan, M. Levin, P. Karlsson and K. Helou (2014). "Frequent MYC coamplification and DNA hypomethylation of multiple genes on 8q in 8p11-p12-amplified breast carcinomas." *Oncogenesis* 3(3): e95-e95.
363. Parsa, A. B., A. Movahedi, H. Taghipour, S. Derrible and A. Mohammadian (2020). "Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis." *Accident Analysis & Prevention* 136: 105405.
364. Paterson, E. K. and S. A. Courtneidge (2018). "Invadosomes are coming: new insights into function and disease relevance." *The FEBS Journal* 285(1): 8-27.
365. Patil, R. H., M. Naveen Kumar, K. M. Kiran Kumar, R. Nagesh, K. Kavya, R. L. Babu, G. T. Ramesh and S. Chidananda Sharma (2018). "Dexamethasone inhibits inflammatory response via down regulation of AP-1 transcription factor in human lung epithelial cells." *Gene* 645: 85-94.

366. Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay (2011). "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825 - 2830.
367. Penke, L. R. and M. Peters-Golden (2019). "Molecular determinants of mesenchymal cell activation in fibroproliferative diseases." *Cellular and Molecular Life Sciences* 76(21): 4179-4201.
368. Perugorria, M. J., L. B. Murphy, N. Fullard, J. B. Chakraborty, D. Vyrla, C. L. Wilson, F. Oakley, J. Mann and D. A. Mann (2013). "Tumor progression locus 2/Cot is required for activation of extracellular regulated kinase in liver injury and toll-like receptor-induced TIMP-1 gene transcription in hepatic stellate cells in mice." *Hepatology* 57(3): 1238-1249.
369. Peyruchaud, O., L. Saier and R. Leblanc (2020) "Autotaxin Implication in Cancer Metastasis and Autoimmune Disorders: Functional Implication of Binding Autotaxin to the Cell Surface." *Cancers* 12 DOI: 10.3390/cancers12010105.
370. Peyser, R., S. MacDonnell, Y. Gao, L. Cheng, Y. Kim, T. Kaplan, Q. Ruan, Y. Wei, M. Ni, C. Adler, W. Zhang, K. Devalaraja-Narashimha, J. Grindley, G. Halasz and L. Morton (2019). "Defining the Activated Fibroblast Population in Lung Fibrosis Using Single-Cell Sequencing." *American Journal of Respiratory Cell and Molecular Biology* 61(1): 74-85.
371. Pezoulas, V. C., C. Papaloukas, M. Veyssiere, A. Goules, A. G. Tzioufas, V. Soumelis and D. I. Fotiadis (2021). "A computational workflow for the detection of candidate diagnostic biomarkers of Kawasaki disease using time-series gene expression data." *Computational and Structural Biotechnology Journal* 19: 3058-3068.
372. Picotto, G., L. R. Morse, N. Nguyen, J. Saltzman and R. Battaglini (2019). "TMEM176A and TMEM176B Are Candidate Regulators of Inhibition of Dendritic Cell Maturation and Function after Chronic Spinal Cord Injury." *Journal of Neurotrauma* 37(3): 528-533.
373. Pilewski, J. M., L. Liu, A. C. Henry, A. V. Knauer and C. A. Feghali-Bostwick (2005). "Insulin-Like Growth Factor Binding Proteins 3 and 5 Are Overexpressed in Idiopathic Pulmonary Fibrosis and Contribute to Extracellular Matrix Deposition." *The American Journal of Pathology* 166(2): 399-407.
374. Pilling, D., T. Fan, D. Huang, B. Kaul and R. H. Gomer (2009). "Identification of Markers that Distinguish Monocyte-Derived Fibrocytes from Monocytes, Macrophages, and Fibroblasts." *PLOS ONE* 4(10): e7475.
375. Plastira, I., E. Bernhart, L. Joshi, C. N. Koyani, H. Strohmaier, H. Reicher, E. Malle and W. Sattler (2020). "MAPK signaling determines lysophosphatidic acid (LPA)-induced inflammation in microglia." *Journal of Neuroinflammation* 17(1): 127.
376. Podolanczuk, A. J., C. C. Thomson, M. Remy-Jardin, L. Richeldi, F. J. Martinez, M. Kolb and G. Raghu (2023). "Idiopathic Pulmonary Fibrosis: State of the Art for 2023." *European Respiratory Journal*: 2200957.
377. Posey, J. E. and F. C. Gherardini (2000). "Lack of a Role for Iron in the Lyme Disease Pathogen." *Science* 288(5471): 1651-1653.
378. Poveda, J., L. C. Tabara, B. Fernandez-Fernandez, C. Martin-Cleary, A. B. Sanz, R. Selgas, A. Ortiz and M. D. Sanchez-Niño (2013). "TWEAK/Fn14 and Non-Canonical NF-kappaB Signaling in Kidney Disease." *Frontiers in Immunology* 4.
379. Pradere, J.-P., J. Klein, S. Gres, C. Guigne, E. Neau, P. Valet, D. Calise, J. Chun, J.-L. Bascands, J.-S. Saulnier-Blache and J. P. Schanstra (2007). "LPA1 Receptor Activation Promotes Renal Interstitial Fibrosis." *Journal of the American Society of Nephrology* 18(12).
380. Prasse, A., H. Binder, J. C. Schupp, G. Kayser, E. Bargagli, B. Jaeger, M. Hess, S. Rittinghausen, L. Vuga, H. Lynn, S. Violette, B. Jung, K. Quast, B. Vanaudenaerde, Y. Xu, J. M. Hohlfeld, N. Krug, J. D. Herazo-Maya, P. Rottoli, W. A. Wuyts and N. Kaminski (2018). "BAL Cell Gene Expression Is Indicative of Outcome and Airway Basal Cell Involvement in Idiopathic Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 199(5): 622-630.

381. Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies and F. O. Glöckner (2013). "The SILVA ribosomal RNA gene database project: improved data processing and web-based tools." *Nucleic Acids Research* 41(D1): D590-D596.
382. Quintavalle, M., L. Elia, G. Condorelli and S. A. Courtneidge (2010). "MicroRNA control of podosome formation in vascular smooth muscle cells in vivo and in vitro." *Journal of Cell Biology* 189(1): 13-22.
383. Rabinovich, E. I., M. G. Kapetanaki, I. Steinfeld, K. F. Gibson, K. V. Pandit, G. Yu, Z. Yakhini and N. Kaminski (2012). "Global Methylation Patterns in Idiopathic Pulmonary Fibrosis." *PLOS ONE* 7(4): e33770.
384. Raghu, G., H. R. Collard, J. J. Egan, F. J. Martinez, J. Behr, K. K. Brown, T. V. Colby, J.-F. Cordier, K. R. Flaherty, J. A. Lasky, D. A. Lynch, J. H. Ryu, J. J. Swigris, A. U. Wells, J. Ancochea, D. Bouros, C. Carvalho, U. Costabel, M. Ebina, D. M. Hansell, T. Johkoh, D. S. Kim, T. E. King, Y. Kondoh, J. Myers, N. L. Müller, A. G. Nicholson, L. Richeldi, M. Selman, R. F. Dudden, B. S. Griss, S. L. Protzko and H. J. Schünemann (2011). "An Official ATS/ERS/JRS/ALAT Statement: Idiopathic Pulmonary Fibrosis: Evidence-based Guidelines for Diagnosis and Management." *American Journal of Respiratory and Critical Care Medicine* 183(6): 788-824.
385. Raghu, G., K. R. Flaherty, D. J. Lederer, D. A. Lynch, T. V. Colby, J. L. Myers, S. D. Groshong, B. T. Larsen, J. H. Chung, M. P. Steele, S. Benzaquen, K. Calero, A. H. Case, G. J. Criner, S. D. Nathan, N. S. Rai, M. Ramaswamy, L. Hagemeyer, J. R. Davis, U. A. Gauhar, D. G. Pankratz, Y. Choi, J. Huang, P. S. Walsh, H. Neville, L. R. Lofaro, N. M. Barth, G. C. Kennedy, K. K. Brown and F. J. Martinez (2019). "Use of a molecular classifier to identify usual interstitial pneumonia in conventional transbronchial lung biopsy samples: a prospective validation study." *The Lancet Respiratory Medicine* 7(6): 487-496.
386. Raimundo, K., E. Chang, M. S. Broder, K. Alexander, J. Zazzali and J. J. Swigris (2016). "Clinical and economic burden of idiopathic pulmonary fibrosis: a retrospective cohort study." *BMC Pulmonary Medicine* 16(1): 2.
387. Rajkumar, R., K. Konishi, T. J. Richards, D. C. Ishizawar, A. C. Wiechert, N. Kaminski and F. Ahmad (2010). "Genomewide RNA expression profiling in lung identifies distinct signatures in idiopathic pulmonary arterial hypertension and secondary pulmonary hypertension." *American Journal of Physiology-Heart and Circulatory Physiology* 298(4): H1235-H1248.
388. Ramani, K., R. J. Tan, D. Zhou, B. M. Coleman, C. V. Jawale, Y. Liu and P. S. Biswas (2018). "IL-17 Receptor Signaling Negatively Regulates the Development of Tubulointerstitial Fibrosis in the Kidney." *Mediators of Inflammation* 2018: 5103672.
389. Ramírez, G., J. S. Hagood, Y. Sanders, R. Ramírez, C. Becerril, L. Segura, L. Barrera, M. Selman and A. Pardo (2011). "Absence of Thy-1 results in TGF- β induced MMP-9 expression and confers a profibrotic phenotype to human lung fibroblasts." *Laboratory Investigation* 91(8): 1206-1218.
390. Ranganathan, P., C. Jayakumar and G. Ramesh (2013). "Proximal tubule-specific overexpression of netrin-1 suppresses acute kidney injury-induced interstitial fibrosis and glomerulosclerosis through suppression of IL-6/STAT3 signaling." *American Journal of Physiology-Renal Physiology* 304(8): F1054-F1065.
391. Rani, J., A. R. Shah and S. Ramachandran (2015). "pubmed.mineR: An R package with text-mining algorithms to analyse PubMed abstracts." *Journal of Biosciences* 40(4): 671-682.
392. Ren, J., P. Lu, X. Zhou, Y. Liao, X. Liu, J. Li, W. Wang, J. Wang, L. Wen, W. Fu and F. Tang (2022). "Genome-Scale Methylation Analysis of Circulating Cell-Free DNA in Gastric Cancer Patients." *Clinical Chemistry* 68(2): 354-364.
393. Renzoni, E. A., D. J. Abraham, S. Howat, X. Shi-Wen, P. Sestini, G. Bou-Gharios, A. U. Wells, S. Veeraghavan, A. G. Nicholson, C. P. Denton, A. Leask, J. D. Pearson, C. M. Black, K. I. Welsh and R. M. du Bois (2004). "Gene expression profiling reveals novel TGF β targets in adult lung fibroblasts." *Respiratory Research* 5(1): 24.

394. Reyfman, P. A., J. M. Walter, N. Joshi, K. R. Anekalla, A. C. McQuattie-Pimentel, S. Chiu, R. Fernandez, M. Akbarpour, C.-I. Chen, Z. Ren, R. Verma, H. Abdala-Valencia, K. Nam, M. Chi, S. Han, F. J. Gonzalez-Gonzalez, S. Soberanes, S. Watanabe, K. J. N. Williams, A. S. Flozak, T. T. Nicholson, V. K. Morgan, D. R. Winter, M. Hinchcliff, C. L. Hrusch, R. D. Guzy, C. A. Bonham, A. I. Sperling, R. Bag, R. B. Hamanaka, G. M. Mutlu, A. V. Yeldandi, S. A. Marshall, A. Shilatifard, L. A. N. Amaral, H. Perlman, J. I. Sznajder, A. C. Argento, C. T. Gillespie, J. Dematte, M. Jain, B. D. Singer, K. M. Ridge, A. P. Lam, A. Bharat, S. M. Bhorade, C. J. Gottardi, G. R. S. Budinger and A. V. Misharin (2018). "Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 199(12): 1517-1536.
395. Richards, T. J., N. Kaminski, F. Baribaud, S. Flavin, C. Brodmerkel, D. Horowitz, K. Li, J. Choi, L. J. Vuga, K. O. Lindell, M. Klesen, Y. Zhang and K. F. Gibson (2012). "Peripheral Blood Proteins Predict Mortality in Idiopathic Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 185(1): 67-76.
396. Richeldi, L., R. M. du Bois, G. Raghu, A. Azuma, K. K. Brown, U. Costabel, V. Cottin, K. R. Flaherty, D. M. Hansell, Y. Inoue, D. S. Kim, M. Kolb, A. G. Nicholson, P. W. Noble, M. Selman, H. Taniguchi, M. Brun, F. Le Maulf, M. Girard, S. Stowasser, R. Schlenker-Herceg, B. Disse and H. R. Collard (2014). "Efficacy and Safety of Nintedanib in Idiopathic Pulmonary Fibrosis." *New England Journal of Medicine* 370(22): 2071-2082.
397. Riemondy, K. A., N. L. Jansing, P. Jiang, E. F. Redente, A. E. Gillen, R. Fu, A. J. Miller, J. R. Spence, A. N. Gerber, J. R. Hesselberth and R. L. Zemans (2019). "Single-cell RNA sequencing identifies TGF- β as a key regenerative cue following LPS-induced lung injury." *JCI Insight* 4(8).
398. Risso, D., K. Schwartz, G. Sherlock and S. Dudoit (2011). "GC-Content Normalization for RNA-Seq Data." *BMC Bioinformatics* 12(1): 480-480.
399. Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi and G. K. Smyth (2015). "limma powers differential expression analyses for RNA-sequencing and microarray studies." *Nucleic acids research* 43(7): e47-e47.
400. Ritchie, M. E., J. Silver, A. Oshlack, M. Holmes, D. Diyagama, A. Holloway and G. K. Smyth (2007). "A comparison of background correction methods for two-colour microarrays." *Bioinformatics* 23(20): 2700-2707.
401. Rizza, C., N. Leitinger, J. Yue, D. J. Fischer, D. A. Wang, P. T. Shih, H. Lee, G. Tigyi and J. A. Berliner (1999). "Lysophosphatidic acid as a regulator of endothelial/leukocyte interaction." *Lab Invest* 79(10): 1227-1235.
402. Rodriguez, L. R., M. Emblom-Callahan, M. Chhina, S. Bui, B. Aljeberry, L. H. Tran, R. Novak, M. Lemma, S. D. Nathan and G. M. Grant (2018). "Global Gene Expression Analysis in an in vitro Fibroblast Model of Idiopathic Pulmonary Fibrosis Reveals Potential Role for CXCL14/CXCR4." *Scientific Reports* 8(1): 3983.
403. Rosas, I. O., T. J. Richards, K. Konishi, Y. Zhang, K. Gibson, A. E. Lokshin, K. O. Lindell, J. Cisneros, S. D. MacDonald, A. Pardo, F. Sciarba, J. Dauber, M. Selman, B. R. Gochuico and N. Kaminski (2008). "MMP1 and MMP7 as Potential Peripheral Blood Biomarkers in Idiopathic Pulmonary Fibrosis." *PLOS Medicine* 5(4): e93.
404. Roulis, M., A. Kaklamanos, M. Scherthanner, P. Bielecki, J. Zhao, E. Kaffe, L.-S. Frommelt, R. Qu, M. S. Knapp, A. Henriques, N. Chalkidi, V. Koliaraki, J. Jiao, J. R. Brewer, M. Bacher, H. N. Blackburn, X. Zhao, R. M. Breyer, V. Aidinis, D. Jain, B. Su, H. R. Herschman, Y. Kluger, G. Kollias and R. A. Flavell (2020). "Paracrine orchestration of intestinal tumorigenesis by a mesenchymal niche." *Nature* 580(7804): 524-529.
405. Roulis, M., C. Nikolaou, E. Kotsaki, E. Kaffe, N. Karagianni, V. Koliaraki, K. Salpea, J. Ragoussis, V. Aidinis, E. Martini, C. Becker, H. R. Herschman, S. Vetrano, S. Danese and G. Kollias (2014). "Intestinal myofibroblast-specific Tpl2-Cox-2-PGE2 pathway links innate sensing to epithelial homeostasis." *Proceedings of the National Academy of Sciences* 111(43): E4658-E4667.

406. Rumelhart, D. E., G. E. Hinton and R. J. Williams (1986). Learning internal representations by error propagation. *Parallel distributed processing: explorations in the microstructure of cognition*, vol. 1: foundations, MIT Press: 318–362.
407. Rustici, G., N. Kolesnikov, M. Brandizi, T. Burdett, M. Dylag, I. Emam, A. Farne, E. Hastings, J. Ison, M. Keays, N. Kurbatova, J. Malone, R. Mani, A. Mupo, R. Pedro Pereira, E. Pilicheva, J. Rung, A. Sharma, Y. A. Tang, T. Ternent, A. Tikhonov, D. Welter, E. Williams, A. Brazma, H. Parkinson and U. Sarkans (2013). "ArrayExpress update—trends in database growth and links to data analysis tools." *Nucleic Acids Research* 41(D1): D987-D990.
408. Rustici, G., E. Williams, M. Barzine, A. Brazma, R. Bumgarner, M. Chierici, C. Furlanello, L. Greger, G. Jurman, M. Miller, F. B. F. Ouellette, J. Quackenbush, M. Reich, C. J. J. Stoeckert, R. C. Taylor, S. C. Trutane, J. Weller, B. Wilhelm and N. Winegarten (2021). "Transcriptomics data availability and reusability in the transition from microarray to next-generation sequencing." *bioRxiv*: 2020.2012.2031.425022.
409. Saatian, B., Y. Zhao, D. He, Steve N. Georas, T. Watkins, Ernst W. Spannhake and V. Natarajan (2006). "Transcriptional regulation of lysophosphatidic acid-induced interleukin-8 expression and secretion by p38 MAPK and JNK in human bronchial epithelial cells." *Biochemical Journal* 393(3): 657-668.
410. Sakai, N., J. Chun, J. S. Duffield, D. Lagares, T. Wada, A. D. Luster and A. M. Tager (2017). "Lysophosphatidic acid signaling through its receptor initiates profibrotic epithelial cell fibroblast communication mediated by epithelial cell derived connective tissue growth factor." *Kidney International* 91(3): 628-641.
411. Sanz, A. B., M. C. Izquierdo, M. D. Sanchez-Niño, A. C. Uceró, J. Egido, M. Ruiz-Ortega, A. M. Ramos, C. Putterman and A. Ortiz (2014). "TWEAK and the progression of renal disease: clinical translation." *Nephrology Dialysis Transplantation* 29(suppl_1): i54-i62.
412. Sanz, A. B., P. Justo, M. D. Sanchez-Niño, L. M. Blanco-Colio, J. A. Winkles, M. Kretzler, A. Jakubowski, J. Blanco, J. Egido, M. Ruiz-Ortega and A. Ortiz (2008). "The Cytokine TWEAK Modulates Renal Tubulointerstitial Inflammation." *Journal of the American Society of Nephrology* 19(4).
413. Savin, I. A., M. A. Zenkova and A. V. Sen'kova (2022) "Pulmonary Fibrosis as a Result of Acute Lung Inflammation: Molecular Mechanisms, Relevant In Vivo Models, Prognostic and Therapeutic Approaches." *International Journal of Molecular Sciences* 23 DOI: 10.3390/ijms232314959.
414. Schafer, M. J., T. A. White, K. Iijima, A. J. Haak, G. Ligresti, E. J. Atkinson, A. L. Oberg, J. Birch, H. Salmonowicz, Y. Zhu, D. L. Mazula, R. W. Brooks, H. Fuhrmann-Stroissnigg, T. Pirtskhalava, Y. S. Prakash, T. Tchkonja, P. D. Robbins, M. C. Aubry, J. F. Passos, J. L. Kirkland, D. J. Tschumperlin, H. Kita and N. K. LeBrasseur (2017). "Cellular senescence mediates fibrotic pulmonary disease." *Nature Communications* 8(1): 14532.
415. Scheraga, R. G., C. Thompson, M. E. Tulapurkar, A. C. Nagarsekar, M. Cowan, R. Potla, J. Sun, R. Cai, C. Logun, J. Shelhamer, N. W. Todd, I. S. Singh, I. G. Luzina, S. P. Atamas and J. D. Hasday (2016). "Activation of heat shock response augments fibroblast growth factor-1 expression in wounded lung epithelium." *American Journal of Physiology-Lung Cellular and Molecular Physiology* 311(5): L941-L955.
416. Schübeler, D. (2015). "Function and information content of DNA methylation." *Nature* 517(7534): 321-326.
417. Seals, D. F., E. F. Azucena, Jr., I. Pass, L. Tesfay, R. Gordon, M. Woodrow, J. H. Resau and S. A. Courtneidge (2005). "The adaptor protein Tks5/Fish is required for podosome formation and function, and for the protease-driven invasion of cancer cells." *Cancer Cell* 7(2): 155-165.
418. Seelan, R. S., P. Mukhopadhyay, M. M. Pisano and R. M. Greene (2018). "Effects of 5-Aza-2'-deoxycytidine (decitabine) on gene expression." *Drug Metabolism Reviews* 50(2): 193-207.
419. Seibold, M. A., A. L. Wise, M. C. Speer, M. P. Steele, K. K. Brown, J. E. Loyd, T. E. Fingerlin, W. Zhang, G. Gudmundsson, S. D. Groshong, C. M. Evans, S. Garantziotis, K. B. Adler, B. F. Dickey, R. M. du Bois, I. V. Yang, A. Herron, D. Kervitsky, J. L. Talbert, C. Markin, J. Park, A. L.

- Crews, S. H. Slifer, S. Auerbach, M. G. Roy, J. Lin, C. E. Hennessy, M. I. Schwarz and D. A. Schwartz (2011). "A Common MUC5B Promoter Polymorphism and Pulmonary Fibrosis." *New England Journal of Medicine* 364(16): 1503-1512.
420. Selman, M., A. Pardo, L. Barrera, A. Estrada, S. R. Watson, K. Wilson, N. Aziz, N. Kaminski and A. Zlotnik (2006). "Gene Expression Profiles Distinguish Idiopathic Pulmonary Fibrosis from Hypersensitivity Pneumonitis." *American Journal of Respiratory and Critical Care Medicine* 173(2): 188-198.
421. Shannon, C. E. (1948). "A Mathematical Theory of Communication." *Bell System Technical Journal* 27(3): 379-423.
422. Shao, X., J. Liao, C. Li, X. Lu, J. Cheng and X. Fan (2021). "CellTalkDB: a manually curated database of ligand–receptor interactions in humans and mice." *Briefings in Bioinformatics* 22(4): bbaa269.
423. Shendure, J., S. Balasubramanian, G. M. Church, W. Gilbert, J. Rogers, J. A. Schloss and R. H. Waterston (2017). "DNA sequencing at 40: past, present and future." *Nature* 550(7676): 345-353.
424. Shi, W., A. Oshlack and G. K. Smyth (2010). "Optimizing the noise versus bias trade-off for Illumina whole genome expression BeadChips." *Nucleic Acids Research* 38(22): e204-e204.
425. Shi, Y., B. R. Gochuico, G. Yu, X. Tang, J. C. Osorio, I. E. Fernandez, C. F. Riquez, A. S. Patel, Y. Shi, M. G. Wathelet, A. J. Goodwin, J. A. Haspel, S. W. Ryter, E. M. Billings, N. Kaminski, D. Morse and I. O. Rosas (2013). "Syndecan-2 Exerts Antifibrotic Effects by Promoting Caveolin-1–mediated Transforming Growth Factor- β Receptor I Internalization and Inhibiting Transforming Growth Factor- β 1 Signaling." *American Journal of Respiratory and Critical Care Medicine* 188(7): 831-841.
426. Shinbori, T., M. Matsuki, M. Suga, K. Kakimoto and M. Ando (1996). "Induction of Interstitial Pneumonia in Autoimmune Mice by Intratracheal Administration of Superantigen Staphylococcal Enterotoxin B." *Cellular Immunology* 174(2): 129-137.
427. Shlyonsky, V., R. Naeije and F. Mies (2014). "Possible Role of Lysophosphatidic Acid in Rat Model of Hypoxic Pulmonary Vascular Remodeling." *Pulmonary Circulation* 4(3): 471-481.
428. Simões, F. C., T. J. Cahill, A. Kenyon, D. Gavriouchkina, J. M. Vieira, X. Sun, D. Pezzolla, C. Ravaut, E. Masmanian, M. Weinberger, S. Mayes, M. E. Lemieux, D. N. Barnette, M. Gunadasa-Rohling, R. M. Williams, D. R. Greaves, L. A. Trinh, S. E. Fraser, S. L. Dallas, R. P. Choudhury, T. Sauka-Spengler and P. R. Riley (2020). "Macrophages directly contribute collagen to scar formation during zebrafish heart regeneration and mouse heart repair." *Nature Communications* 11(1): 600.
429. Simone, S., C. Cosola, A. Loverre, M. Cariello, F. Sallustio, F. Rascio, L. Gesualdo, F. P. Schena, G. Grandaliano and G. Pertosa (2012). "BMP-2 induces a profibrotic phenotype in adult renal progenitor cells through Nox4 activation." *American Journal of Physiology-Renal Physiology* 303(1): F23-F34.
430. Sindhu, S., N. Akhter, A. Wilson, R. Thomas, H. Arefanian, A. Al Madhoun, F. Al-Mulla and R. Ahmad (2020) "MIP-1 α Expression Induced by Co-Stimulation of Human Monocytic Cells with Palmitate and TNF- α Involves the TLR4-IRF3 Pathway and Is Amplified by Oxidative Stress." *Cells* 9 DOI: 10.3390/cells9081799.
431. Sivakumar, P., J. R. Thompson, R. Ammar, M. Porteous, C. McCoubrey, E. Cantu, K. Ravi, Y. Zhang, Y. Luo, D. Streltsov, M. F. Beers, G. Jarai and J. D. Christie (2019). "RNA sequencing of transplant-stage idiopathic pulmonary fibrosis lung reveals unique pathway regulation." *ERJ Open Research* 5(3): 00117-02019.
432. Smith, L. C., S. Moreno, L. Robertson, S. Robinson, K. Gant, A. J. Bryant and T. Sabo-Attwood (2018). "Transforming growth factor beta1 targets estrogen receptor signaling in bronchial epithelial cells." *Respiratory Research* 19(1): 160.
433. Smyth, G. K. and T. Speed (2003). "Normalization of cDNA microarray data." *Methods* 31(4): 265-273.

434. Spinardi, L., J. Rietdorf, L. Nitsch, M. Bono, C. Tacchetti, M. Way and P. C. Marchisio (2004). "A dynamic podosome-like structure of epithelial cells." *Experimental Cell Research* 295(2): 360-374.
435. Staab-Weijnitz, C. A., I. E. Fernandez, L. Knüppel, J. Maul, K. Heinzemann, B. M. Juan-Guardela, E. Hennen, G. Preissler, H. Winter, C. Neurohr, R. Hatz, M. Lindner, J. Behr, N. Kaminski and O. Eickelberg (2015). "FK506-Binding Protein 10, a Potential Novel Drug Target for Idiopathic Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 192(4): 455-467.
436. Stangou, M., A. Papagianni, C. Bantis, D. Moisiadis, S. Kasimatis, M. Spartalis, A. Pantzaki, G. Efstratiadis and D. Memmos (2013). "Up-regulation of urinary markers predict outcome in IgA nephropathy but their predictive value is influenced by treatment with steroids and azathioprine." *Clin Nephrol* 80(3): 203-210.
437. Stanislawski, M. A., D. Dabelea, L. A. Lange, B. D. Wagner and C. A. Lozupone (2019). "Gut microbiota phenotypes of obesity." *npj Biofilms and Microbiomes* 5(1): 18.
438. Stefanovic, B. (2013). "RNA protein interactions governing expression of the most abundant protein in human body, type I collagen." *WIREs RNA* 4(5): 535-545.
439. Sterne, J. A. C., S. Murthy, J. V. Diaz, A. S. Slutsky, J. Villar, D. C. Angus, D. Annane, L. C. P. Azevedo, O. Berwanger, A. B. Cavalcanti, P.-F. Dequin, B. Du, J. Emberson, D. Fisher, B. Giraudeau, A. C. Gordon, A. Granholm, C. Green, R. Haynes, N. Heming, J. P. T. Higgins, P. Horby, P. Juni, M. J. Landray, A. Le Gouge, M. Leclerc, W. S. Lim, F. R. Machado, C. McArthur, F. Mezzani, M. H. Moller, A. Perner, M. W. Petersen, J. Savovic, B. Tomazini, V. C. Veiga, S. Webb and J. C. Marshall (2020). "Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19: A Meta-analysis." *JAMA* 324(13): 1330-1341.
440. Stoddard, S. F., B. J. Smith, R. Hein, B. R. K. Roller and T. M. Schmidt (2015). "rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development." *Nucleic Acids Research* 43(D1): D593-D598.
441. Strunz, M., L. M. Simon, M. Ansari, J. J. Kathiriya, I. Angelidis, C. H. Mayr, G. Tsidiridis, M. Lange, L. F. Mattner, M. Yee, P. Ogar, A. Sengupta, I. Kukhtevich, R. Schneider, Z. Zhao, C. Voss, T. Stoeger, J. H. L. Neumann, A. Hilgendorff, J. Behr, M. O'Reilly, M. Lehmann, G. Burgstaller, M. Königshoff, H. A. Chapman, F. J. Theis and H. B. Schiller (2020). "Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis." *Nature Communications* 11(1): 3559.
442. Stuart, B. D., J. Choi, S. Zaidi, C. Xing, B. Holohan, R. Chen, M. Choi, P. Dharwadkar, F. Torres, C. E. Girod, J. Weissler, J. Fitzgerald, C. Kershaw, J. Klesney-Tait, Y. Mageto, J. W. Shay, W. Ji, K. Bilguvar, S. Mane, R. P. Lifton and C. K. Garcia (2015). "Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening." *Nature Genetics* 47(5): 512-517.
443. Stuart, T., A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, III, Y. Hao, M. Stoeckius, P. Smibert and R. Satija (2019). "Comprehensive Integration of Single-Cell Data." *Cell* 177(7): 1888-1902.e1821.
444. Subramanian, A., R. Narayan, S. M. Corsello, D. D. Peck, T. E. Natoli, X. Lu, J. Gould, J. F. Davis, A. A. Tubelli, J. K. Asiedu, D. L. Lahr, J. E. Hirschman, Z. Liu, M. Donahue, B. Julian, M. Khan, D. Wadden, I. C. Smith, D. Lam, A. Liberzon, C. Toder, M. Bagul, M. Orzechowski, O. M. Enache, F. Piccioni, S. A. Johnson, N. J. Lyons, A. H. Berger, A. F. Shamji, A. N. Brooks, A. Vrcic, C. Flynn, J. Rosains, D. Y. Takeda, R. Hu, D. Davison, J. Lamb, K. Ardlie, L. Hogstrom, P. Greenside, N. S. Gray, P. A. Clemons, S. Silver, X. Wu, W.-N. Zhao, W. Read-Button, X. Wu, S. J. Haggarty, L. V. Ronco, J. S. Boehm, S. L. Schreiber, J. G. Doench, J. A. Bittker, D. E. Root, B. Wong and T. R. Golub (2017). "A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles." *Cell* 171(6): 1437-1452.e1417.
445. Sun, B., H. Wang, L. Zhang, X. Yang, M. Zhang, X. Zhu, X. Ji and H. Wang (2018). "Role of interleukin 17 in TGF- β signaling-mediated renal interstitial fibrosis." *Cytokine* 106: 80-88.

446. Sun, S., R. Wang, J. Song, M. Guan, N. Li, X. Zhang, Z. Zhao and J. Zhang (2017). "Blocking gp130 signaling suppresses autotaxin expression in adipocytes and improves insulin sensitivity in diet-induced obesity." *Journal of Lipid Research* 58(11): 2102-2113.
447. Sung, H., J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray (2021). "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries." *CA: A Cancer Journal for Clinicians* 71(3): 209-249.
448. Suppli, M. P., J. I. Bagger, B. Lelouvier, A. Broha, M. Demant, M. J. Kønig, C. Strandberg, A. Lund, T. Vilsbøll and F. K. Knop (2021). "Hepatic microbiome in healthy lean and obese humans." *JHEP Reports* 3(4).
449. Svensson, V., R. Vento-Tormo and S. A. Teichmann (2018). "Exponential scaling of single-cell RNA-seq in the past decade." *Nature Protocols* 13(4): 599-604.
450. Swaney, J. S., C. Chapman, L. D. Correa, K. J. Stebbins, A. R. Broadhead, G. Bain, A. M. Santini, J. Darlington, C. D. King, C. S. Baccei, C. Lee, T. A. Parr, J. R. Roppe, T. J. Seiders, J. Ziff, P. Prasit, J. H. Hutchinson, J. F. Evans and D. S. Lorrain (2011). "Pharmacokinetic and Pharmacodynamic Characterization of an Oral Lysophosphatidic Acid Type 1 Receptor-Selective Antagonist." *Journal of Pharmacology and Experimental Therapeutics* 336(3): 693.
451. Szklarczyk, D., A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen and Christian v. Mering (2019). "STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets." *Nucleic Acids Research* 47(D1): D607-D613.
452. Tabib, T., M. Huang, N. Morse, A. Papazoglou, R. Behera, M. Jia, M. Bulik, D. E. Monier, P. V. Benos, W. Chen, R. Domsic and R. Lafyatis (2021). "Myofibroblast transcriptome indicates SFRP2hi fibroblast progenitors in systemic sclerosis skin." *Nature Communications* 12(1): 4384.
453. Takeda, Y., K. Matoba, D. Kawanami, Y. Nagai, T. Akamine, S. Ishizawa, Y. Kanazawa, T. Yokota and K. Utsunomiya (2019) "ROCK2 Regulates Monocyte Migration and Cell to Cell Adhesion in Vascular Endothelial Cells." *International Journal of Molecular Sciences* 20 DOI: 10.3390/ijms20061331.
454. Tan, J., J. R. Tedrow, J. A. Dutta, B. Juan-Guardela, M. Nouraie, Y. Chu, H. Trejo Bittar, K. Ramani, P. S. Biswas, K. L. Veraldi, N. Kaminski, Y. Zhang and D. J. Kass (2016). "Expression of RXFP1 Is Decreased in Idiopathic Pulmonary Fibrosis. Implications for Relaxin-based Therapies." *American Journal of Respiratory and Critical Care Medicine* 194(11): 1392-1402.
455. Tang, F., C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch, A. Siddiqui, K. Lao and M. A. Surani (2009). "mRNA-Seq whole-transcriptome analysis of a single cell." *Nature Methods* 6(5): 377-382.
456. Tashiro, J., G. A. Rubio, A. H. Limper, K. Williams, S. J. Elliot, I. Ninou, V. Aidinis, A. Tzouveleakis and M. K. Glassberg (2017). "Exploring Animal Models That Resemble Idiopathic Pulmonary Fibrosis." *Frontiers in Medicine* 4.
457. Tate, P. H. and A. P. Bird (1993). "Effects of DNA methylation on DNA-binding proteins and gene expression." *Current Opinion in Genetics & Development* 3(2): 226-231.
458. Tazi, A., F. Bouchonnet, M. Grandsaigne, L. Boumsell, A. J. Hance and P. Soler (1993). "Evidence that granulocyte macrophage-colony-stimulating factor regulates the distribution and differentiated state of dendritic cells/Langerhans cells in human lung and lung cancers." *The Journal of Clinical Investigation* 91(2): 566-576.
459. The Recovery Collaborative Group (2020). "Dexamethasone in Hospitalized Patients with Covid-19." *New England Journal of Medicine* 384(8): 693-704.
460. Tomalka, J. A., A. N. Pelletier, S. Fourati, M. B. Latif, A. Sharma, K. Furr, K. Carlson, M. Lifton, A. Gonzalez, P. Wilkinson, G. Franchini, R. Parks, N. Letvin, N. Yates, K. Seaton, G. Tomaras, J. Tartaglia, M. L. Robb, N. L. Michael, R. Koup, B. Haynes, S. Santra and R. P. Sekaly (2021). "The transcription factor CREB1 is a mechanistic driver of immunogenicity and reduced HIV-1 acquisition following ALVAC vaccination." *Nature Immunology* 22(10): 1294-1305.

461. Tomasek, J. J., G. Gabbiani, B. Hinz, C. Chaponnier and R. A. Brown (2002). "Myofibroblasts and mechano-regulation of connective tissue remodelling." *Nature Reviews Molecular Cell Biology* 3(5): 349-363.
462. Tong, X., F. Su, X. Xu, H. Xu, T. Yang, Q. Xu, H. Dai, K. Huang, L. Zou, W. Zhang, S. Pei, F. Xiao, Y. Li and C. Wang (2019). "Alterations to the Lung Microbiome in Idiopathic Pulmonary Fibrosis Patients." *Frontiers in Cellular and Infection Microbiology* 9.
463. Travaglini, K. J., A. N. Nabhan, L. Penland, R. Sinha, A. Gillich, R. V. Sit, S. Chang, S. D. Conley, Y. Mori, J. Seita, G. J. Berry, J. B. Shrager, R. J. Metzger, C. S. Kuo, N. Neff, I. L. Weissman, S. R. Quake and M. A. Krasnow (2020). "A molecular cell atlas of the human lung from single-cell RNA sequencing." *Nature* 587(7835): 619-625.
464. Treekitkarnmongkol, W., M. Hassane, A. Sinjab, K. Chang, K. Hara, Z. Rahal, J. Zhang, W. Lu, S. Sivakumar, T. L. McDowell, J. Kantrowitz, J. Zhou, W. Lang, L. Xu, J. K. Ochieng, S. Nunomura-Nakamura, S. Deng, C. Behrens, M. G. Raso, J. Fukuoka, A. Reuben, E. J. Ostrin, E. Parra, L. M. Solis, A. E. Spira, F. McAllister, T. Cascone, I. I. Wistuba, S. J. Moghaddam, P. A. Scheet, J. Fujimoto and H. Kadara (2020). "Augmented Lipocalin-2 Is Associated with Chronic Obstructive Pulmonary Disease and Counteracts Lung Adenocarcinoma Development." *American Journal of Respiratory and Critical Care Medicine* 203(1): 90-101.
465. Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Board on Health Care Services, Board on Health Sciences Policy, Institute of Medicine (2012). *Omics-Based Clinical Discovery: Science, Technology, and Applications. Evolution of Translational Omics: Lessons Learned and the Path Forward*. C. Micheel, S. Nass and G. Omenn. <https://www.ncbi.nlm.nih.gov/books/NBK202165/>, National Academies Press (US).
466. Triebel, S., J. Bläser, H. Reinke and H. Tschesche (1992). "A 25 kDa α 2-microglobulin-related protein is a component of the 125 kDa form of human gelatinase." *FEBS Letters* 314(3): 386-388.
467. Tripathi, A., J. Debelius, D. A. Brenner, M. Karin, R. Loomba, B. Schnabl and R. Knight (2018). "The gut–liver axis and the intersection with the microbiome." *Nature Reviews Gastroenterology & Hepatology* 15(7): 397-411.
468. Tsitoura, E., E. Vasarmidi, E. Bibaki, A. Trachalaki, C. Koutoulaki, G. Papastratigakis, S. Papadogiorgaki, G. Chalepakis, N. Tzanakis and K. M. Antoniou (2019). "Accumulation of damaged mitochondria in alveolar macrophages with reduced OXPHOS related gene expression in IPF." *Respiratory Research* 20(1): 264.
469. Tsitoura, E., A. U. Wells, K. Karagiannis, I. Lasithiotaki, E. Vasarmidi, E. Bibaki, C. Koutoulaki, H. Sato, D. A. Spandidos, N. M. Sifakas, G. Sourvinos and K. M. Antoniou (2016). "MiR-185/AKT and miR-29a/Collagen 1a pathways are activated in IPF BAL cells." *Oncotarget*; Vol 7, No 46.
470. Tsukui, T., K.-H. Sun, J. B. Wetter, J. R. Wilson-Kanamori, L. A. Hazelwood, N. C. Henderson, T. S. Adams, J. C. Schupp, S. D. Poli, I. O. Rosas, N. Kaminski, M. A. Matthay, P. J. Wolters and D. Sheppard (2020). "Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis." *Nature Communications* 11(1): 1920.
471. Tunçay, S. C., E. Doğan, G. Hakverdi, Z. Ü. Tutar and S. Mir (2021). "Interleukin-8 is increased in chronic kidney disease in children, but not related to cardiovascular disease." *Brazilian Journal of Nephrology* 43.
472. Tzouvelekis, A., V. Harokopos, T. Paparountas, N. Oikonomou, A. Chatziioannou, G. Vilaras, E. Tsiambas, A. Karameris, D. Bouros and V. Aidinis (2007). "Comparative Expression Profiling in Pulmonary Fibrosis Suggests a Role of Hypoxia-inducible Factor-1 α in Disease Pathogenesis." *American Journal of Respiratory and Critical Care Medicine* 176(11): 1108-1119.
473. Tzouvelekis, A., P. Ntoliou, A. Karameris, G. Vilaras, P. Boglou, A. Koulelidis, K. Archontogeorgis, K. Kaltsas, G. Zacharis, E. Sarikloglou, P. Steiropoulos, D. Mikroulis, A. Koutsopoulos, M. Froudarakis and D. Bouros (2013). "Increased Expression of Epidermal Growth Factor Receptor (EGF-R) in Patients with Different Forms of Lung Fibrosis." *BioMed Research International* 2013: 654354.

474. Tzouveleakis, A., G. Yu, C. L. Lino Cardenas, J. D. Herazo-Maya, R. Wang, T. Woolard, Y. Zhang, K. Sakamoto, H. Lee, J.-S. Yi, G. DeIuliis, N. Xylourgidis, F. Ahangari, P. J. Lee, V. Aidinis, E. L. Herzog, R. Homer, A. M. Bennett and N. Kaminski (2016). "SH2 Domain-Containing Phosphatase-2 Is a Novel Antifibrotic Regulator in Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 195(4): 500-514.
475. Uceró, A. C., A. Benito-Martin, I. Fuentes-Calvo, B. Santamaria, J. Blanco, J. M. Lopez-Novoa, M. Ruiz-Ortega, J. Egido, L. C. Burkly, C. Martinez-Salgado and A. Ortiz (2013). "TNF-related weak inducer of apoptosis (TWEAK) promotes kidney fibrosis and Ras-dependent proliferation of cultured renal fibroblast." *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1832(10): 1744-1755.
476. Udomsinprasert, W., N. Kitkumthorn, A. Mutirangura, V. Chongsrisawat, Y. Poovorawan and S. Honsawek (2017). "Association between Promoter Hypomethylation and Overexpression of Autotaxin with Outcome Parameters in Biliary Atresia." *PLOS ONE* 12(1): e0169306.
477. Ulhaq, Z. S. and G. V. Soraya (2020). "Interleukin-6 as a potential biomarker of COVID-19 progression." *Médecine et Maladies Infectieuses* 50(4): 382-383.
478. Urushiyama, H., T. Jo, W. Hasegawa, A. Yokoyama, T. Ando, Y. Sakamoto, R. Kumazawa, K. Uda, N. Michihata, N. Awano, M. Hiroki, K. Fushimi, H. Yasunaga and T. Nagase (2022). "Effect of nintedanib on acute exacerbations of fibrosing interstitial lung diseases: a national database study in Japan." *ERJ Open Research* 8(4): 00209-02022.
479. Valenzi, E., M. Bulik, T. Tabib, C. Morse, J. Sembrat, H. T. Bittar, M. Rojas and R. Lafyatis (2019). "Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung disease." *Annals of the Rheumatic Diseases* 78(10): 1379.
480. Valenzi, E., T. Tabib, A. Papazoglou, J. Sembrat, H. E. Trejo Bittar, M. Rojas and R. Lafyatis (2021). "Disparate Interferon Signaling and Shared Aberrant Basaloid Cells in Single-Cell Profiling of Idiopathic Pulmonary Fibrosis and Systemic Sclerosis-Associated Interstitial Lung Disease." *Frontiers in Immunology* 12.
481. Van Der Maaten, L. and G. Hinton (2008). "Visualizing Data using t-SNE." *Journal of Machine Learning Research* 9(86): 2579-2605.
482. Vancheri, C. (2013). "Common pathways in idiopathic pulmonary fibrosis and cancer." *European Respiratory Review* 22(129): 265.
483. Vassiliou, A. G., C. Keskinidou, E. Jahaj, P. Gallos, I. Dimopoulou, A. Kotanidou and S. E. Orfanos (2021) "ICU Admission Levels of Endothelial Biomarkers as Predictors of Mortality in Critically Ill COVID-19 Patients." *Cells* 10 DOI: 10.3390/cells10010186.
484. Vazirinejad, R., Z. Ahmadi, M. Kazemi Arababadi, G. Hassanshahi and D. Kennedy (2014). "The Biological Functions, Structure and Sources of CXCL10 and Its Outstanding Part in the Pathophysiology of Multiple Sclerosis." *Neuroimmunomodulation* 21(6): 322-330.
485. Větrovský, T. and P. Baldrian (2013). "The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses." *PLOS ONE* 8(2): e57923.
486. Viedt, C. and S. R. Orth (2002). "Monocyte chemoattractant protein-1 (MCP-1) in the kidney: does it more than simply attract monocytes?" *Nephrology Dialysis Transplantation* 17(12): 2043-2047.
487. Vielhauer, V., H.-J. Anders, M. Mack, J. Cihak, F. Strutz, M. Stangassinger, B. Luckow, H.-J. Grone and D. Schlöndorff (2001). "Obstructive Nephropathy in the Mouse: Progressive Fibrosis Correlates with Tubulointerstitial Chemokine Expression and Accumulation of CC Chemokine Receptor 2- and 5-Positive Leukocytes." *Journal of the American Society of Nephrology* 12(6).
488. Vielhauer, V. and T. N. Mayadas (2007). "Functions of TNF and its Receptors in Renal Disease: Distinct Roles in Inflammatory Tissue Injury and Immune Regulation." *Seminars in Nephrology* 27(3): 286-308.
489. Villaseñor-Altamirano, A. B., M. Moretto, M. Maldonado, A. Zayas-Del Moral, A. Munguía-Reyes, Y. Romero, J. S. García-Sotelo, L. A. Aguilar, O. Aldana-Assad, K. Engelen, M. Selman, J.

- Collado-Vides, Y. I. Balderas-Martínez and A. Medina-Rivera (2020). "PulmonDB: a curated lung disease gene expression database." *Scientific Reports* 10(1): 514.
490. Vougioukalaki, M., D. C. Kanellis, K. Gkouskou and A. G. Eliopoulos (2011). "Tpl2 kinase signal transduction in inflammation and cancer." *Cancer Letters* 304(2): 80-89.
491. Vukmirovic, M., J. D. Herazo-Maya, J. Blackmon, V. Skodric-Trifunovic, D. Jovanovic, S. Pavlovic, J. Stojsic, V. Zeljkovic, X. Yan, R. Homer, B. Stefanovic and N. Kaminski (2017). "Identification and validation of differentially expressed transcripts by RNA-sequencing of formalin-fixed, paraffin-embedded (FFPE) lung tissue from patients with Idiopathic Pulmonary Fibrosis." *BMC Pulmonary Medicine* 17(1): 15.
492. Vukmirovic, M. and N. Kaminski (2018). "Impact of Transcriptomics on Our Understanding of Pulmonary Fibrosis." *Frontiers in Medicine* 5: 87.
493. Wada, T., H. Yokoyama, K. Furuichi, K.-I. Kobayashi, K. Harada, M. Naruto, S.-B. Su, M. Akiyama, N. Mukaida and K. Matsushima (1996). "Intervention of crescentic glomerulonephritis by antibodies to monocyte chemotactic and activating factor (MCAF/MCP-1)." *The FASEB Journal* 10(12): 1418-1425.
494. Walker, E., A. V. Hernandez and M. W. Kattan (2008). "Meta-analysis: Its strengths and limitations." *Cleveland Clinic Journal of Medicine* 75(6): 431.
495. Walsh, S. L. F., L. Calandriello, M. Silva and N. Sverzellati (2018). "Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study." *The Lancet Respiratory Medicine* 6(11): 837-845.
496. Wan, H., X. Huang, P. Cong, M. He, A. Chen, T. Wu, D. Dai, W. Li, X. Gao, L. Tian, H. Liang and L. Xiong (2021). "Identification of Hub Genes and Pathways Associated With Idiopathic Pulmonary Fibrosis via Bioinformatics Analysis." *Frontiers in Molecular Biosciences* 8.
497. Wang, B., A. Law, T. Regan, N. Parkinson, J. Cole, C. D. Russell, D. H. Dockrell, M. U. Gutmann and J. K. Baillie (2022). "Systematic comparison of ranking aggregation methods for gene lists in experimental results." *Bioinformatics*: btac621.
498. Wang, D., B. Deng, L. Cheng, J. Li, J. Zhang, X. Zhang, X. Guo, T. Yan, X. Yue, Y. An, B. Zhang, W. Yang, J. Xie and R. Wang (2023). "A novel and low-toxic peptide DR3penA alleviates pulmonary fibrosis by regulating the MAPK/miR-23b-5p/AQP5 signaling axis." *Acta Pharmaceutica Sinica B* 13(2): 722-738.
499. Wang, J., M. Jiang, A. Xiong, L. Zhang, L. Luo, Y. Liu, S. Liu, Q. Ran, D. Wu, Y. Xiong, X. He, E. L.-H. Leung and G. Li (2022). "Integrated analysis of single-cell and bulk RNA sequencing reveals pro-fibrotic PLA2G7high macrophages in pulmonary fibrosis." *Pharmacological Research* 182: 106286.
500. Wang, S.-C., L.-M. Liao, M. Ansar, S.-Y. Lin, W.-W. Hsu, C.-M. Su, Y.-M. Chung, C.-C. Liu, C.-S. Hung and R.-K. Lin (2021) "Automatic Detection of the Circulating Cell-Free Methylated DNA Pattern of GCM2, ITPRIPL1 and CCDC181 for Detection of Early Breast Cancer and Surgical Treatment Response." *Cancers* 13 DOI: 10.3390/cancers13061375.
501. Wang, X. M., Y. Zhang, H. P. Kim, Z. Zhou, C. A. Feghali-Bostwick, F. Liu, E. Ifedigbo, X. Xu, T. D. Oury, N. Kaminski and A. M. K. Choi (2006). "Caveolin-1: a critical regulator of lung fibrosis in idiopathic pulmonary fibrosis." *The Journal of Experimental Medicine* 203(13): 2895.
502. Wang, Y., J. Yella, J. Chen, F. X. McCormack, S. K. Madala and A. G. Jegga (2017). "Unsupervised gene expression analyses identify IPF-severity correlated signatures, associated genes and biomarkers." *BMC Pulmonary Medicine* 17(1): 133.
503. Wang, Z., J. Zhu, F. Chen and L. Ma (2019). "Weighted Gene Coexpression Network Analysis Identifies Key Genes and Pathways Associated with Idiopathic Pulmonary Fibrosis." *Medical Science Monitor* 25: 4285-4304.
504. Warszawska, J. M., R. Gawish, O. Sharif, S. Sigel, B. Doninger, K. Lakovits, I. Mesteri, M. Nairz, L. Boon, A. Spiel, V. Fuhrmann, B. Strobl, M. Müller, P. Schenk, G. Weiss and S. Knapp (2013).

- "Lipocalin 2 deactivates macrophages and worsens pneumococcal pneumonia outcomes." *The Journal of Clinical Investigation* 123(8): 3363-3372.
505. Wasnick, R. M., M. Korfei, K. Piskulak, I. Henneke, J. Wilhelm, P. Mahavadi, D. von der Beck, M. Koch, I. Shalashova, O. Klymenko, L. Fink, H. Witt, H. Hackstein, E. El Agha, S. Bellusci, W. Klepetko, M. Königshoff, O. Eickelberg, T. Braun, W. Seeger, C. Ruppert and A. Guenther (2019). "Restored alveolar epithelial differentiation and reversed human lung fibrosis upon Notch inhibition." *bioRxiv*: 580498.
506. Watford, W. T., C.-C. Wang, C. Tsatsanis, L. A. Mielke, A. G. Eliopoulos, C. Daskalakis, N. Charles, S. Odom, J. Rivera, J. O'Shea and P. N. Tsichlis (2009). "Ablation of Tumor Progression Locus 2 Promotes a Type 2 Th Cell Response in Ovalbumin-Immunized Mice." *The Journal of Immunology* 184(1): 105-113.
507. Weinberg, E. D. (1997). "The Lactobacillus Anomaly: Total Iron Abstinence." *Perspectives in Biology and Medicine* 40(4): 578-583.
508. Wells, A. U., K. R. Flaherty, K. K. Brown, Y. Inoue, A. Devaraj, L. Richeldi, T. Moua, B. Crestani, W. A. Wuyts, S. Stowasser, M. Quaresma, R.-G. Goeldner, R. Schlenker-Herceg, M. Kolb, S. Abe, M. Aburto, O. Acosta, C. Andrews, D. Antin-Ozerkis, G. Arce, M. Arias, S. Avdeev, A. Barczyk, R. Bascom, E. Bazdyrev, P. Beirne, E. Belloli, M. A. Bergna, E. Bergot, N. Bhatt, S. Blaas, B. Bondue, F. Bonella, E. Britt, K. Buch, J. Burk, H. Cai, A. Cantin, D. M. Castillo Villegas, A. Cazaux, S. Cerri, S. Chaaban, N. Chaudhuri, V. Cottin, B. Crestani, G. Criner, C. Dahlqvist, S. Danoff, J. Dematte D'Amico, D. Dilling, P. Elias, N. Ettinger, J. Falk, E. R. Fernández Pérez, A. Gamez-Dubuis, G. Giessel, A. Gifford, M. Glassberg, C. Glazer, J. Golden, L. Gómez Carrera, J. Guiot, R. Hallowell, H. Hayashi, J. Hetzel, N. Hirani, L. Homik, B. Hope-Gill, D. Hotchkin, K. Ichikado, M. Ilkovich, Y. Inoue, S. Izumi, E. Jassem, L. Jones, S. Jouneau, R. Kaner, J. Kang, T. Kawamura, R. Kessler, Y. Kim, K. Kishi, H. Kitamura, M. Kolb, Y. Kondoh, C. Kono, D. Koschel, M. Kreuter, T. Kulkarni, J. Kus, F. Lebargy, A. León Jiménez, Q. Luo, Y. Mageto, T. M. Maher, S. Makino, S. Marchand-Adam, C. Marquette, R. Martinez, M. Martínez, R. Maturana Rozas, Y. Miyazaki, S. Moiseev, M. Molina-Molina, L. Morrison, L. Morrow, T. Moua, A. Nambiar, Y. Nishioka, H. Nunes, M. Okamoto, J. Oldham, M. Otaola, M. Padilla, J. S. Park, N. Patel, A. Pesci, W. Piotrowski, L. Pitts, H. Poonyagariyagorn, A. Prasse, S. Quadrelli, W. Randerath, R. Refini, M. Reynaud-Gaubert, F. Riviere, J. A. Rodríguez Portal, I. Rosas, M. Rossman, Z. Safdar, T. Saito, N. Sakamoto, M. Salinas Fénero, J. Sauleda, S. Schmidt, M. B. Scholand, M. Schwartz, S. Shapera, O. Shlobin, B. Sigal, A. Silva Orellana, D. Skowasch, J. W. Song, S. Stieglitz, H. Stone, M. Streck, T. Suda, H. Sugiura, H. Takahashi, H. Takaya, T. Takeuchi, K. Thavarajah, L. Tolle, S. Tomassetti, K. Tomii, C. Valenzuela, C. Vancheri, F. Varone, S. Veeraraghavan, A. Villar, S. Weigt, L. Wemeau, W. Wuyts, Z. Xu, V. Yakusevich, Y. Yamada, H. Yamauchi and D. Ziora (2020). "Nintedanib in patients with progressive fibrosing interstitial lung diseases; subgroup analyses by interstitial lung disease diagnosis in the INBUILD trial: a randomised, double-blind, placebo-controlled, parallel-group trial." *The Lancet Respiratory Medicine* 8(5): 453-460.
509. Wen, A. Y., K. M. Sakamoto and L. S. Miller (2010). "The Role of the Transcription Factor CREB in Immune Function." *The Journal of Immunology* 185(11): 6413-6419.
510. White, E. S. (2015). "Lung Extracellular Matrix and Fibroblast Function." *Annals of the American Thoracic Society* 12(Supplement 1): S30-S33.
511. White, E. S., V. J. Thannickal, S. L. Carskadon, E. G. Dickie, D. L. Livant, S. Markwart, G. B. Toews and D. A. Arenberg (2003). "Integrin $\alpha 4 \beta 1$ Regulates Migration across Basement Membranes by Lung Fibroblasts." *American Journal of Respiratory and Critical Care Medicine* 168(4): 436-442.
512. Wilborn, J., L. J. Crofford, M. D. Burdick, S. L. Kunkel, R. M. Strieter and M. Peters-Golden (1995). "Cultured lung fibroblasts isolated from patients with idiopathic pulmonary fibrosis have a diminished capacity to synthesize prostaglandin E2 and to express cyclooxygenase-2." *The Journal of Clinical Investigation* 95(4): 1861-1868.
513. Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P.

- Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons (2016). "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3(1): 160018.
514. Wingett, S. W. and S. Andrews (2018). "FastQ Screen: A tool for multi-genome mapping and quality control." *F1000 Research*.
515. Winter, D. J. (2017). "rentrez: An R package for the NCBI eUtils API." *The R Journal* 9(2): 520-526.
516. Wu, H., Y. Yu, H. Huang, Y. Hu, S. Fu, Z. Wang, M. Shi, X. Zhao, J. Yuan, J. Li, X. Yang, E. Bin, D. Wei, H. Zhang, J. Zhang, C. Yang, T. Cai, H. Dai, J. Chen and N. Tang (2020). "Progressive Pulmonary Fibrosis Is Caused by Elevated Mechanical Tension on Alveolar Stem Cells." *Cell* 180(1): 107-121.e117.
517. Wu, J.-M., Y. Xu, N. J. Skill, H. Sheng, Z. Zhao, M. Yu, R. Saxena and M. A. Maluccio (2010). "Autotaxin expression and its connection with the TNF-alpha-NF-κB axis in human hepatocellular carcinoma." *Molecular Cancer* 9(1): 71.
518. Wu, Z., R. A. Irizarry, R. Gentleman, F. Martinez-Murillo and F. Spencer (2004). "A Model-Based Background Adjustment for Oligonucleotide Expression Arrays." *Journal of the American Statistical Association* 99(468): 909-917.
519. Wynes, M. W., B. L. Edelman, A. G. Kostyk, M. G. Edwards, C. Coldren, S. D. Groshong, G. P. Cosgrove, E. F. Redente, A. Bamberg, K. K. Brown, N. Reisdorph, R. C. Keith, S. K. Frankel and D. W. H. Riches (2011). "Increased Cell Surface Fas Expression Is Necessary and Sufficient To Sensitize Lung Fibroblasts to Fas Ligation-Induced Apoptosis: Implications for Fibroblast Accumulation in Idiopathic Pulmonary Fibrosis." *The Journal of Immunology* 187(1): 527.
520. Xi, Y., T. Kim, A. N. Brumwell, I. H. Driver, Y. Wei, V. Tan, J. R. Jackson, J. Xu, D.-K. Lee, J. E. Gotts, M. A. Matthay, J. M. Shannon, H. A. Chapman and A. E. Vaughan (2017). "Local lung hypoxia determines epithelial fate decisions during alveolar regeneration." *Nature Cell Biology* 19(8): 904-914.
521. Xia, H., V. Bodempudi, A. Benyumov, P. Hergert, D. Tank, J. Herrera, J. Braziunas, O. Larsson, M. Parker, D. Rossi, K. Smith, M. Peterson, A. Limper, J. Jessurun, J. Connett, D. Ingbar, S. Phan, P. B. Bitterman and C. A. Henke (2014). "Identification of a Cell-of-Origin for Fibroblasts Comprising the Fibrotic Reticulum in Idiopathic Pulmonary Fibrosis." *The American Journal of Pathology* 184(5): 1369-1383.
522. Xia, H., A. Gilbertsen, J. Herrera, E. Racila, K. Smith, M. Peterson, T. Griffin, A. Benyumov, L. Yang, P. B. Bitterman and C. A. Henke (2017). "Calcium-binding protein S100A4 confers mesenchymal progenitor cell fibrogenicity in idiopathic pulmonary fibrosis." *The Journal of Clinical Investigation* 127(7): 2586-2597.
523. Xiao, X., B. S. Yeoh and M. Vijay-Kumar (2017). "Lipocalin 2: An Emerging Player in Iron Homeostasis and Inflammation." *Annual Review of Nutrition* 37(1): 103-130.
524. Xie, T., V. Kulur, N. Liu, N. Deng, Y. Wang, S. C. Rowan, C. Yao, G. Huang, X. Liu, F. Taghavifar, J. Liang, C. Hogaboam, B. Stripp, P. Chen, D. Jiang and P. W. Noble (2021). "Mesenchymal growth hormone receptor deficiency leads to failure of alveolar progenitor cell function and severe pulmonary fibrosis." *Science Advances* 7(24): eabg6005.
525. Xie, T., Y. Wang, N. Deng, G. Huang, F. Taghavifar, Y. Geng, N. Liu, V. Kulur, C. Yao, P. Chen, Z. Liu, B. Stripp, J. Tang, J. Liang, P. W. Noble and D. Jiang (2018). "Single-Cell Deconvolution of Fibroblast Heterogeneity in Mouse Pulmonary Fibrosis." *Cell Reports* 22(13): 3625-3640.
526. Xiucheng, L., Q. Xichun, Q. Hao, J. Caili, Y. Yanliang, S. Teng, C. Bi, C. Chang and Z. Hao (2021). "Characterization of the heterogeneity of endothelial cells in bleomycin-induced lung fibrosis using single-cell RNA sequencing." *Angiogenesis* 24(4): 809-821.

527. Xu, T., J. Qiao, L. Zhao, G. He, K. Li, J. Wang, Y. Tian and H. Wang (2009). "Effect of dexamethasone on acute respiratory distress syndrome induced by the H5N1 virus in mice." *European Respiratory Journal* 33(4): 852.
528. Xu, Y., T. Mizuno, A. Sridharan, Y. Du, M. Guo, J. Tang, K. A. Wikenheiser-Brokamp, A.-K. T. Perl, V. A. Funari, J. J. Gokey, B. R. Stripp and J. A. Whitsett (2017). "Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis." *JCI Insight* 1(20).
529. Xu, Z., L. Mo, X. Feng, M. Huang and L. Li (2020). "Using bioinformatics approach identifies key genes and pathways in idiopathic pulmonary fibrosis." *Medicine* 99(36).
530. Yamashita, C. M., L. Dolgonos, R. L. Zemans, S. K. Young, J. Robertson, N. Briones, T. Suzuki, M. N. Campbell, J. Gauldie, D. C. Radisky, D. W. H. Riches, G. Yu, N. Kaminski, C. A. G. McCulloch and G. P. Downey (2011). "Matrix Metalloproteinase 3 Is a Mediator of Pulmonary Fibrosis." *The American Journal of Pathology* 179(4): 1733-1745.
531. Yamashita, S., A. Maeshima, I. Kojima and Y. Nojima (2004). "Activin A Is a Potent Activator of Renal Interstitial Fibroblasts." *Journal of the American Society of Nephrology* 15(1).
532. Yanagihara, T., S. Sato, C. Upagupta and M. Kolb (2019). "What have we learned from basic science studies on idiopathic pulmonary fibrosis?" *European Respiratory Review* 28(153): 190029.
533. Yang, I. V., L. H. Burch, M. P. Steele, J. D. Savov, J. W. Hollingsworth, E. McElvania-Tekippe, K. G. Berman, M. C. Speer, T. A. Sporn, K. K. Brown, M. I. Schwarz and D. A. Schwartz (2007). "Gene Expression Profiling of Familial and Sporadic Interstitial Pneumonia." *American Journal of Respiratory and Critical Care Medicine* 175(1): 45-54.
534. Yang, I. V., C. D. Coldren, S. M. Leach, M. A. Seibold, E. Murphy, J. Lin, R. Rosen, A. J. Neidermyer, D. F. McKean, S. D. Groshong, C. Cool, G. P. Cosgrove, D. A. Lynch, K. K. Brown, M. I. Schwarz, T. E. Fingerlin and D. A. Schwartz (2013). "Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis." *Thorax* 68(12): 1114.
535. Yang, I. V., L. G. Luna, J. Cotter, J. Talbert, S. M. Leach, R. Kidd, J. Turner, N. Kummer, D. Kervitsky, K. K. Brown, K. Boon, M. I. Schwarz, D. A. Schwartz and M. P. Steele (2012). "The Peripheral Blood Transcriptome Identifies the Presence and Extent of Disease in Idiopathic Pulmonary Fibrosis." *PLOS ONE* 7(6): e37708.
536. Yang, S., H. Cui, N. Xie, M. Icyuz, S. Banerjee, V. B. Antony, E. Abraham, V. J. Thannickal and G. Liu (2013). "miR-145 regulates myofibroblast differentiation and lung fibrosis." *The FASEB Journal* 27(6): 2382-2391.
537. Yang, S. Y., J. Lee, C. G. Park, S. Kim, S. Hong, H. C. Chung, S. K. Min, J. W. Han, H. W. Lee and H. Y. Lee (2002). "Expression of autotaxin (NPP-2) is closely linked to invasiveness of breast cancer cells." *Clinical & Experimental Metastasis* 19(7): 603-608.
538. Yang, W., P. C. Rosenstiel and H. Schulenburg (2016). "ABSSeq: A new RNA-Seq analysis method based on modelling absolute expression differences." *BMC Genomics* 17(1): 1-14.
539. Yang, Y. H. J. and N. P. Thorne (2003). Normalization for two-color cDNA microarray data.
540. Yao, C., X. Guan, G. Carraro, T. Parimon, X. Liu, G. Huang, A. Mulay, H. J. Soukiasian, G. David, S. S. Weigt, J. A. Belperio, P. Chen, D. Jiang, P. W. Noble and B. R. Stripp (2020). "Senescence of Alveolar Type 2 Cells Drives Progressive Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 203(6): 707-717.
541. Yates, A. D., P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, A. G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, J. C. Marugán, C. Cummins, C. Davidson, K. Dodiya, R. Fatima, A. Gall, C. G. Giron, L. Gil, T. Grego, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, M. Kay, I. Lavidas, T. Le, D. Lemos, J. G. Martinez, T. Maurel, M. McDowall, A. McMahan, S. Mohanan, B. Moore, M. Nuhn, D. N. Oheh, A. Parker, A. Parton, M. Patricio, M. P. Sakthivel, A. I. Abdul Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, M. Sycheva, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, A. Vullo, B. Walts, A. Winterbottom, A. Zadissa, M. Chakiachvili, B. Flint, A. Frankish, S. E. Hunt, G. Iisley, M. Kostadima, N. Langridge, J. E. Loveland, F. J. Martin, J. Morales, J. M. Mudge, M.

- Muffato, E. Perry, M. Ruffier, S. J. Trevanion, F. Cunningham, K. L. Howe, D. R. Zerbino and P. Flicek (2020). "Ensembl 2020." *Nucleic Acids Research* 48(D1): D682-D688.
542. Ye, Z. and Y. Hu (2021). "TGF- β 1: Gentlemanly orchestrator in idiopathic pulmonary fibrosis (Review)." *Int J Mol Med* 48(1): 132.
543. Young, R. P., R. J. Hopkins and B. Marsland (2015). "The Gut–Liver–Lung Axis. Modulation of the Innate Immune Response and Its Possible Role in Chronic Obstructive Pulmonary Disease." *American Journal of Respiratory Cell and Molecular Biology* 54(2): 161-169.
544. Yu, G., E. Kovkarova-Naumovski, P. Jara, A. Parwani, D. Kass, V. Ruiz, C. Lopez-Otín, I. O. Rosas, K. F. Gibson, S. Cabrera, R. Ramírez, S. A. Yousem, T. J. Richards, L. J. Chensny, M. Selman, N. Kaminski and A. Pardo (2012). "Matrix Metalloproteinase-19 Is a Key Regulator of Lung Fibrosis in Mice and Humans." *American Journal of Respiratory and Critical Care Medicine* 186(8): 752-762.
545. Yu, G., A. Tzouveleakis, R. Wang, J. D. Herazo-Maya, G. H. Ibarra, A. Srivastava, J. P. W. de Castro, G. DeIuliiis, F. Ahangari, T. Woolard, N. Aurelien, R. Arrojo e Drigo, Y. Gan, M. Graham, X. Liu, R. J. Homer, T. S. Scanlan, P. Mannam, P. J. Lee, E. L. Herzog, A. C. Bianco and N. Kaminski (2018). "Thyroid hormone inhibits lung fibrosis in mice by improving epithelial mitochondrial function." *Nature Medicine* 24(1): 39-49.
546. Yu, G., L.-G. Wang, Y. Han and Q.-Y. He (2012). "clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters." *OMICS: A Journal of Integrative Biology* 16(5): 284-287.
547. Yu, G., L.-G. Wang, G.-R. Yan and Q.-Y. He (2015). "DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis." *Bioinformatics* 31(4): 608-609.
548. Yu, X., P. Gu, Z. Huang, X. Fang, Y. Jiang, Q. Luo, X. Li, X. Zhu, M. Zhan, J. Wang, L. Fan, R. Chen, J. Yu, Y. Gu, A. Liang and X. Yi (2017). "Reduced expression of BMP3 contributes to the development of pulmonary fibrosis and predicts the unfavorable prognosis in IIP patients." *Oncotarget*; Vol 8, No 46.
549. Zannikou, M., I. Barbayianni, D. Fanidis, T. Grigorakaki, E. Vlachopoulou, D. Konstantopoulos, M. Fousteri, I. Nikitopoulou, A. Kotanidou, E. Kaffe and V. Aidinis (2021). "MAP3K8 Regulates Cox-2–Mediated Prostaglandin E₂ Production in the Lung and Suppresses Pulmonary Inflammation and Fibrosis." *The Journal of Immunology* 206(3): 607-620.
550. Zeng, Z., Bas G. J. Surewaard, Connie H. Y. Wong, Joan A. Geoghegan, Craig N. Jenne and P. Kubers (2016). "CR1g Functions as a Macrophage Pattern Recognition Receptor to Directly Bind and Capture Blood-Borne Gram-Positive Bacteria." *Cell Host & Microbe* 20(1): 99-106.
551. Zepp, J. A., W. J. Zacharias, D. B. Frank, C. A. Cavanaugh, S. Zhou, M. P. Morley and E. E. Morrisey (2017). "Distinct Mesenchymal Lineages and Niches Promote Epithelial Self-Renewal and Myofibrogenesis in the Lung." *Cell* 170(6): 1134-1148.e1110.
552. Zhang, D., S. Li, N. Wang, H.-Y. Tan, Z. Zhang and Y. Feng (2020). "The Cross-Talk Between Gut Microbiota and Lungs in Common Lung Diseases." *Frontiers in Microbiology* 11: 301.
553. Zhang, D., G. Povysil, C. A. Newton, T. M. Maher, P. L. Molyneaux, I. Noth, F. J. Martinez, G. Raghu, J. L. Todd, S. M. Palmer, A. Platt, S. Petrovski, D. B. Goldstein and C. K. Garcia (2022). "Genome-wide Enrichment of TERT Rare Variants in Idiopathic Pulmonary Fibrosis Patients of Latino Ancestry." *American Journal of Respiratory and Critical Care Medicine* 206(7): 903-905.
554. Zhang, M.-Z., X. Wang, H. Yang, A. B. Fogo, B. J. Murphy, R. Kaltenbach, P. Cheng, B. Zinker and R. C. Harris (2017). "Lysophosphatidic Acid Receptor Antagonism Protects against Diabetic Nephropathy in a Type 2 Diabetic Model." *Journal of the American Society of Nephrology* 28(11).
555. Zhang, Z.-H., H. Chen, N. D. Vaziri, J.-R. Mao, L. Zhang, X. Bai and Y.-Y. Zhao (2016). "Metabolomic Signatures of Chronic Kidney Disease of Diverse Etiologies in the Rats and Humans." *Journal of Proteome Research* 15(10): 3802-3812.

556. Zhang, Z.-L., Y.-L. Hou, D.-T. Li and F.-Z. Li (2020). "Laboratory findings of COVID-19: a systematic review and meta-analysis." *Scandinavian Journal of Clinical and Laboratory Investigation* 80(6): 441-447.
557. Zhao, C., W. Hui, M. J. Fernandes, P. E. Poubelle and S. G. Bourgoïn (2014). "Lysophosphatidic acid-induced IL-8 secretion involves MSK1 and MSK2 mediated activation of CREB1 in human fibroblast-like synoviocytes." *Biochemical Pharmacology* 90(1): 62-72.
558. Zhao, P., Y. Xu, Y. Wei, Q. Qiu, T.-L. Chew, Y. Kang and C. Cheng (2016). "The CD44s splice isoform is a central mediator for invadopodia activity." *Journal of Cell Science* 129(7): 1355-1365.
559. Zhao, Y. D., L. Yin, S. Archer, C. Lu, G. Zhao, Y. Yao, L. Wu, M. Hsin, T. K. Waddell, S. Keshavjee, J. Granton and M. de Perrot (2017). "Metabolic heterogeneity of idiopathic pulmonary fibrosis: a metabolomic study." *BMJ Open Respiratory Research* 4(1): e000183.
560. Zhong, X., Y. J. Tu, Y. Li, P. Zhang, W. Wang, S. S. Chen, L. Li, A. C. Chung, H. Y. Lan, Y. H. Chen, G. S. Li and L. Wang (2017). "Serum levels of WNT1-inducible signaling pathway protein-1 (WISP-1): a noninvasive biomarker of renal fibrosis in subjects with chronic kidney disease." *American Journal of Translational Research* 9(6): 2920-2932.
561. Zhou, R., K. K.-W. To, Y.-C. Wong, L. Liu, B. Zhou, X. Li, H. Huang, Y. Mo, T.-Y. Luk, T. T.-K. Lau, P. Yeung, W.-M. Chan, A. K.-L. Wu, K.-C. Lung, O. T.-Y. Tsang, W.-S. Leung, I. F.-N. Hung, K.-Y. Yuen and Z. Chen (2020). "Acute SARS-CoV-2 Infection Impairs Dendritic Cell and T Cell Responses." *Immunity* 53(4): 864-877.e865.
562. Zhu, K., A. Xu, W. Xia, P. Li, R. Han, E. Wang, S. Zhou and R. Wang (2021). "Integrated analysis of the molecular mechanisms in idiopathic pulmonary fibrosis." *International Journal of Medical Sciences* 18(15): 3412-3424.
563. Zhu, L., S. S. Baker, C. Gill, W. Liu, R. Alkhoury, R. D. Baker and S. R. Gill (2013). "Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH) patients: A connection between endogenous alcohol and NASH." *Hepatology* 57(2): 601-609.
564. Zhu, X., D. Wei, O. Chen, Z. Zhang, J. Xue, S. Huang, W. Zhu and Y. Wang (2016). "Upregulation of CCL3/MIP-1 α regulated by MAPKs and NF-kappaB mediates microglial inflammatory response in LPS-induced brain injury." *Acta Neurobiol Exp (Wars)* 76(4): 304-317.
565. Zhu, Y., S. Davis, R. Stephens, P. S. Meltzer and Y. Chen (2008). "GEOmetadb: powerful alternative search engine for the Gene Expression Omnibus." *Bioinformatics* 24(23): 2798-2800.
566. Zhuang, H., H. Wang and Z. Ji (2022). "findPC: An R package to automatically select the number of principal components in single-cell analysis." *Bioinformatics* 38(10): 2949-2951.
567. Zuo, F., N. Kaminski, E. Eugui, J. Allard, Z. Yakhini, A. Ben-Dor, L. Lollini, D. Morris, Y. Kim, B. DeLustro, D. Sheppard, A. Pardo, M. Selman and R. A. Heller (2002). "Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans." *Proceedings of the National Academy of Sciences* 99(9): 6292.

Appendix I – Supplementary figures

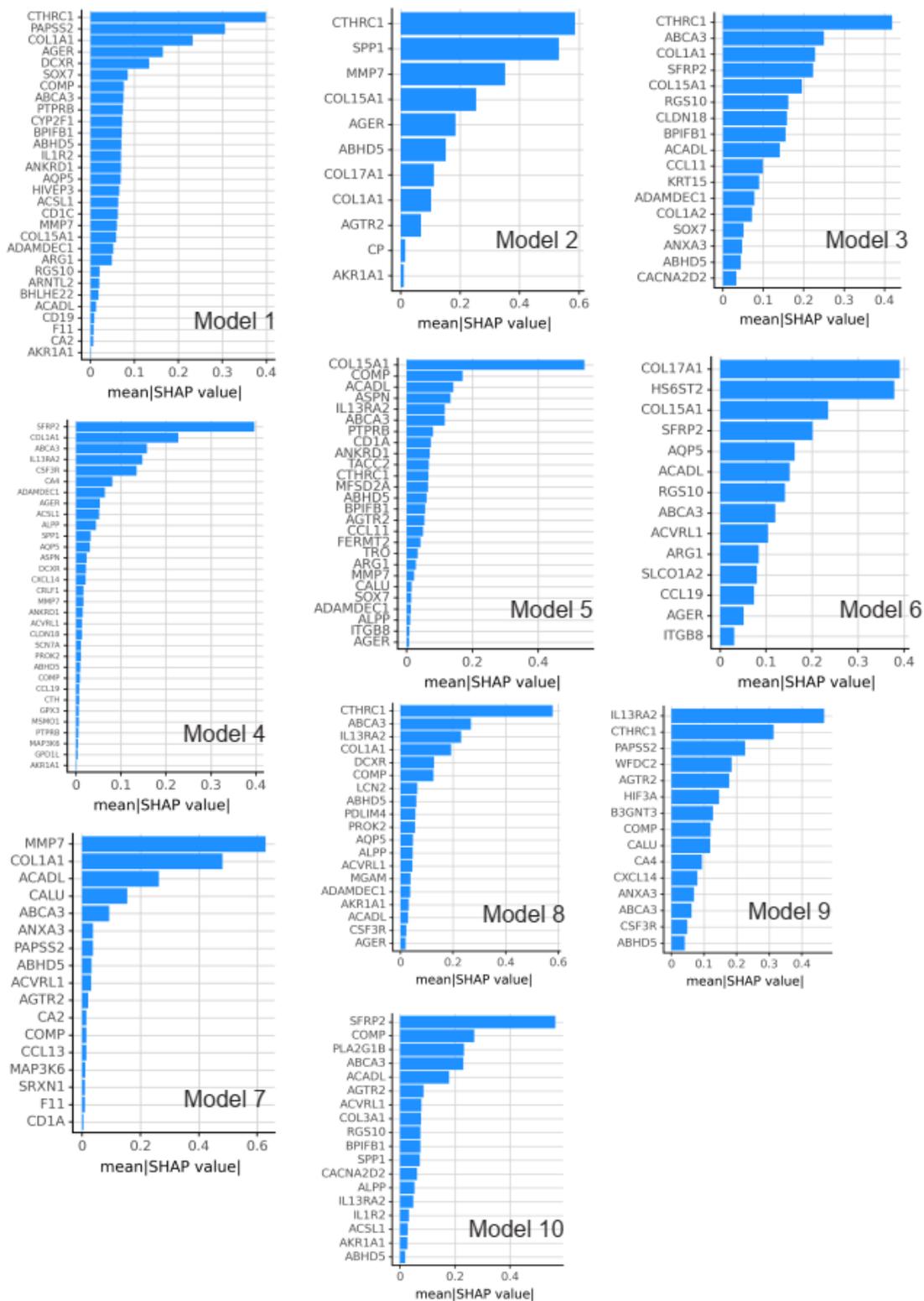


Figure S1. Per model features ranked by mean absolute Shapley (SHAP) value. Phenotypic classification was achieved with different genomic features combination per model. Adopted from (Fanidis, Pezoulas et al. 2023).

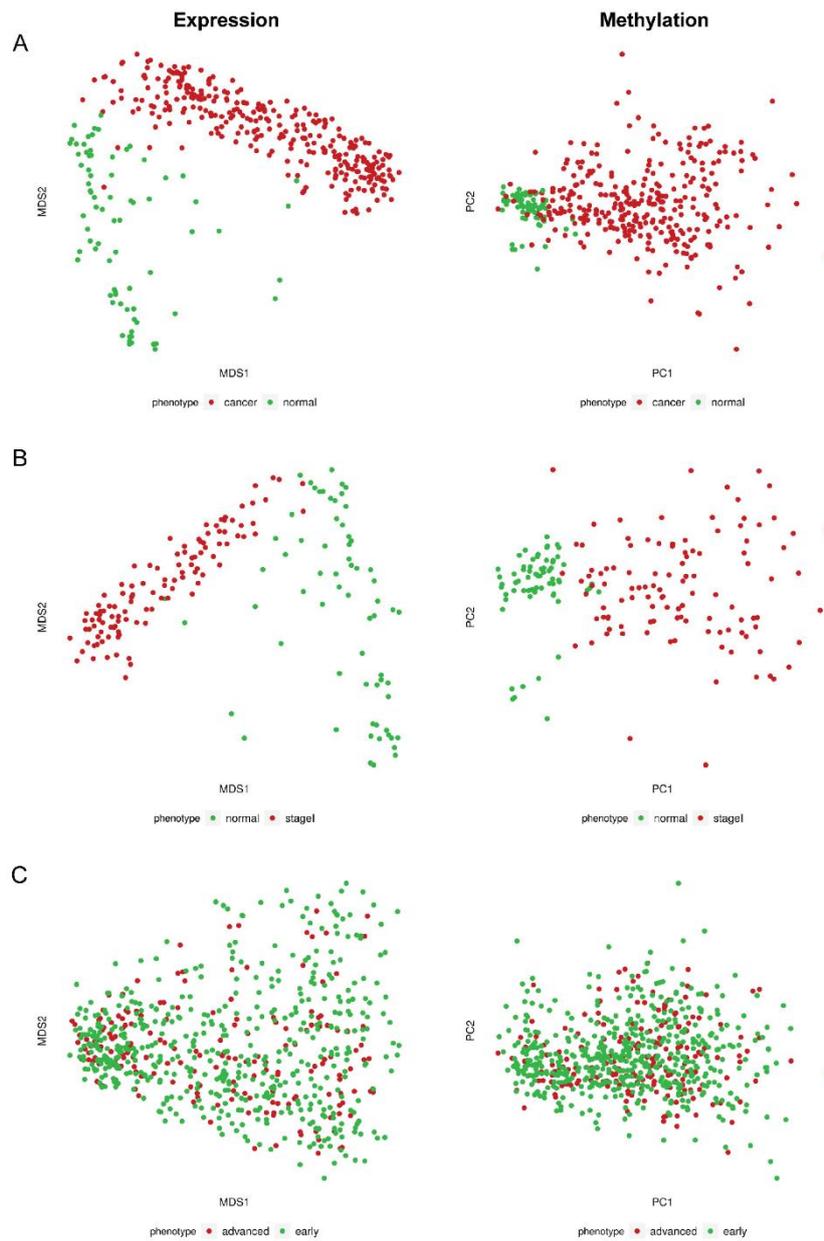


Figure S2. TCGA expression and methylation data projected into reduced space. A. BrCa and normal tissue samples. **B.** Stage I BrCa and normal tissue samples. **C.** Advanced and early BrCa samples. Adopted from (Panagopoulou, Drosouni et al. 2022).

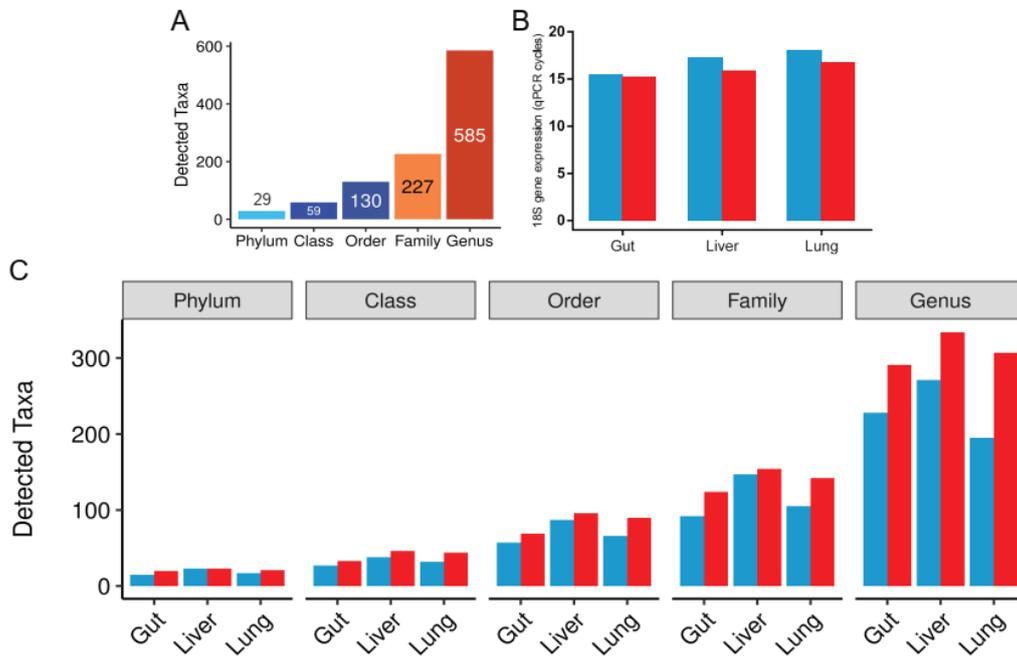


Figure S3. Obesity creates a tendency of increasing microbial wealth in the gut-liver-lung axis. A. Total number of detected taxa. **B.** No changes in host tissue 18S rRNA gene (Q-PCR), ensures equal loading of samples. **C.** Obesity by HFD created a general pattern of increased number of taxa compared to the control diet. Adopted from (Galaris, Fanidis et al. 2022).

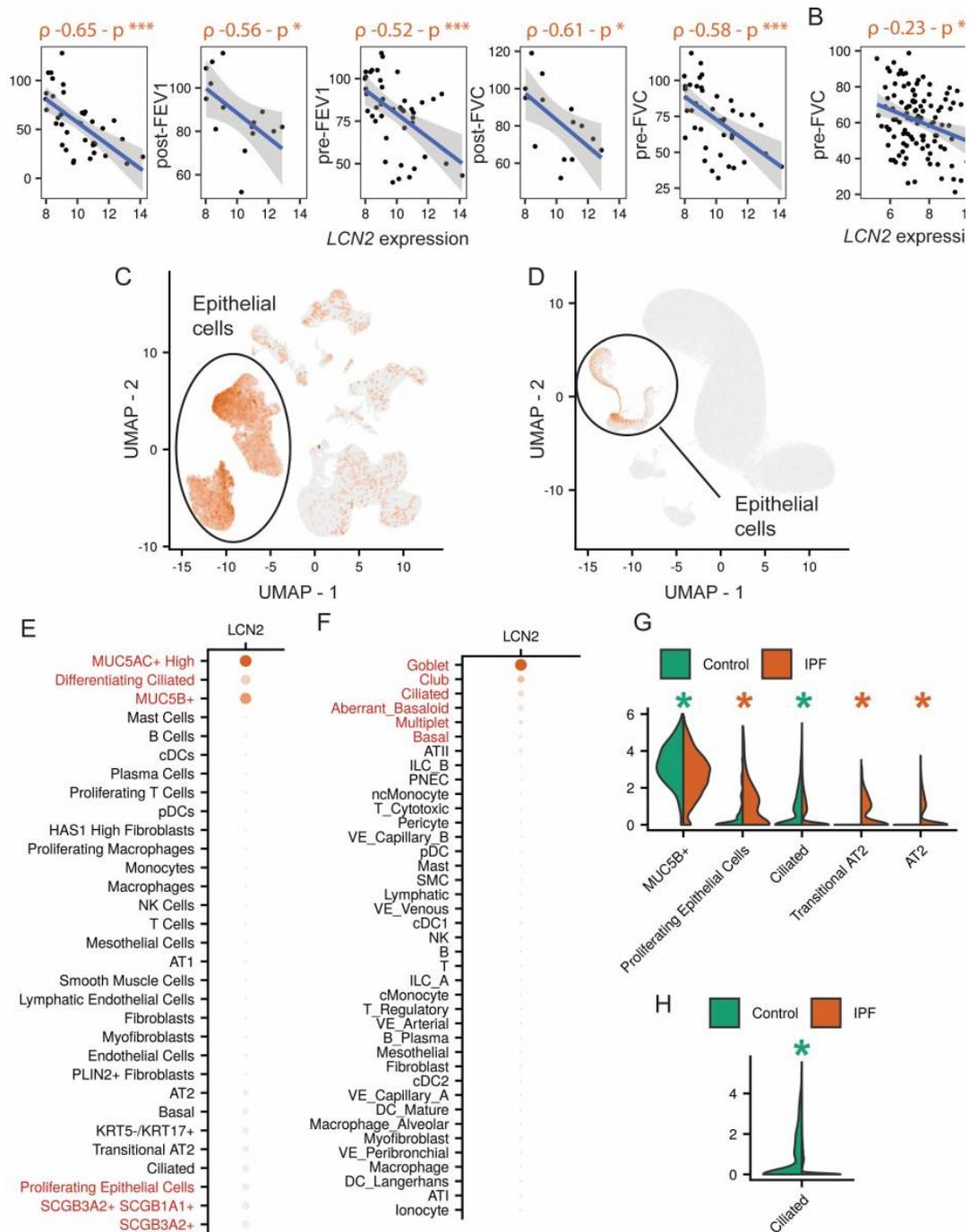


Figure S4. *LCN2* up-regulation in IPF lung tissue is negatively related to lung function. A-B. Correlation between *LCN2* expression and spirometry measurements from two patient cohorts (GSE47460_GPL6480; GSE32537). **C-D.** Dimensionality reduction plots of two single cell datasets (Habermann, Gutierrez et al. 2020; Adams Shupp et al. 2020) that support *LCN2* production from cells of the epithelial lineage. **E-F.** Dot plots of the datasets in C-D, respectively. Red color highlights cell types for which *LCN2* is a marker gene (Wilcoxon tank sum test; FC > 1.2; Bonferroni adjusted p-value < 0.05). **G-H.** Differential expression analysis within cells of the same type and between phenotypes (Wilcoxon tank sum test; |FC| ≥ 1.2; Bonferroni adjusted p-value < 0.05); * up-regulated in IPF; * down-regulated in IPF. Adopted from (Galaris, Fanidis et al. 2023).

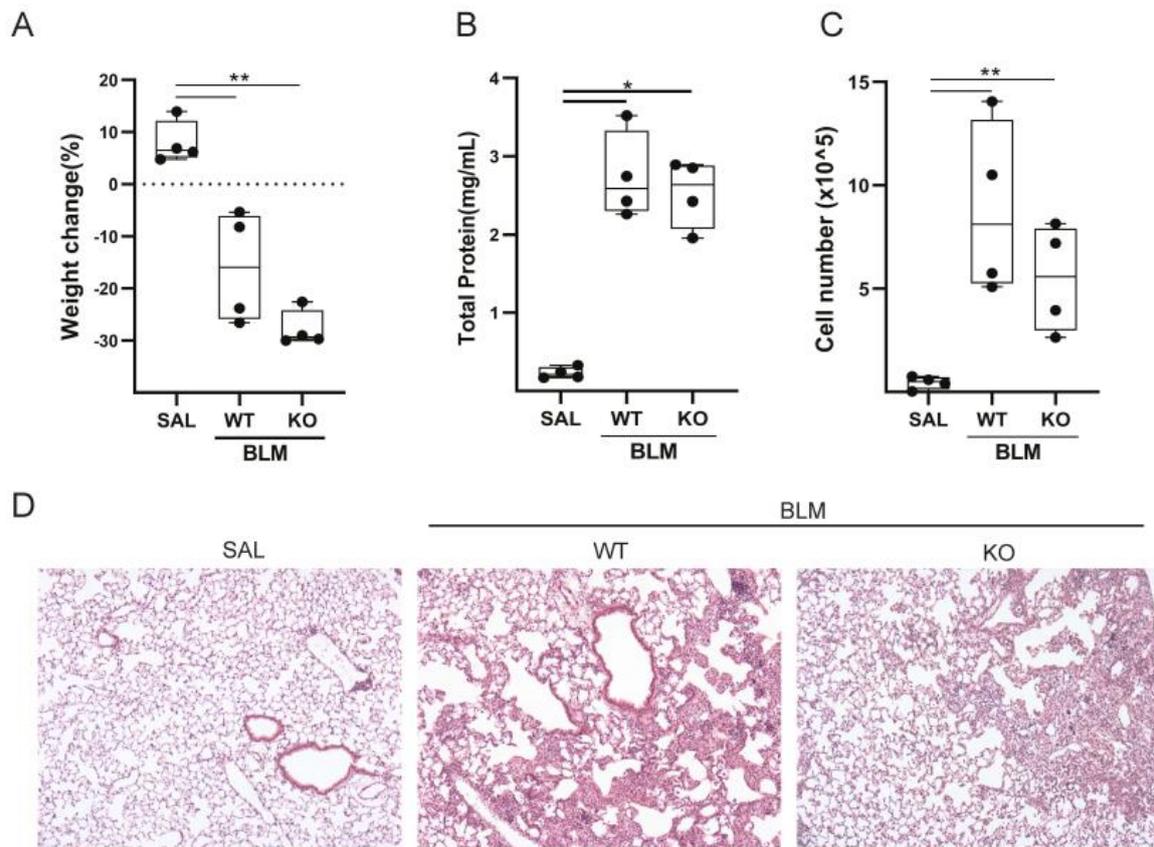


Figure S5. Obese BLM mice models of PF are only slightly impacted by *Lcn2* genetic deficiency. **A.** Change of weight in WT and KO mice treated with BLM compared to their saline-treated counterparts. **B.** BAL fluid protein concentration (Bradford assay). **C.** Number of cells in the BAL fluid of obese mice on the 14th day post BLM administration (hepatocytometer). Statistical significance was evaluated using Kruskal-Wallis test (** $p < 0.01$). **D.** Stainings of murine lung tissue sections from WT and *Lcn2* knockout mice (x10). Adopted from (Galaris, Fanidis et al. 2023).

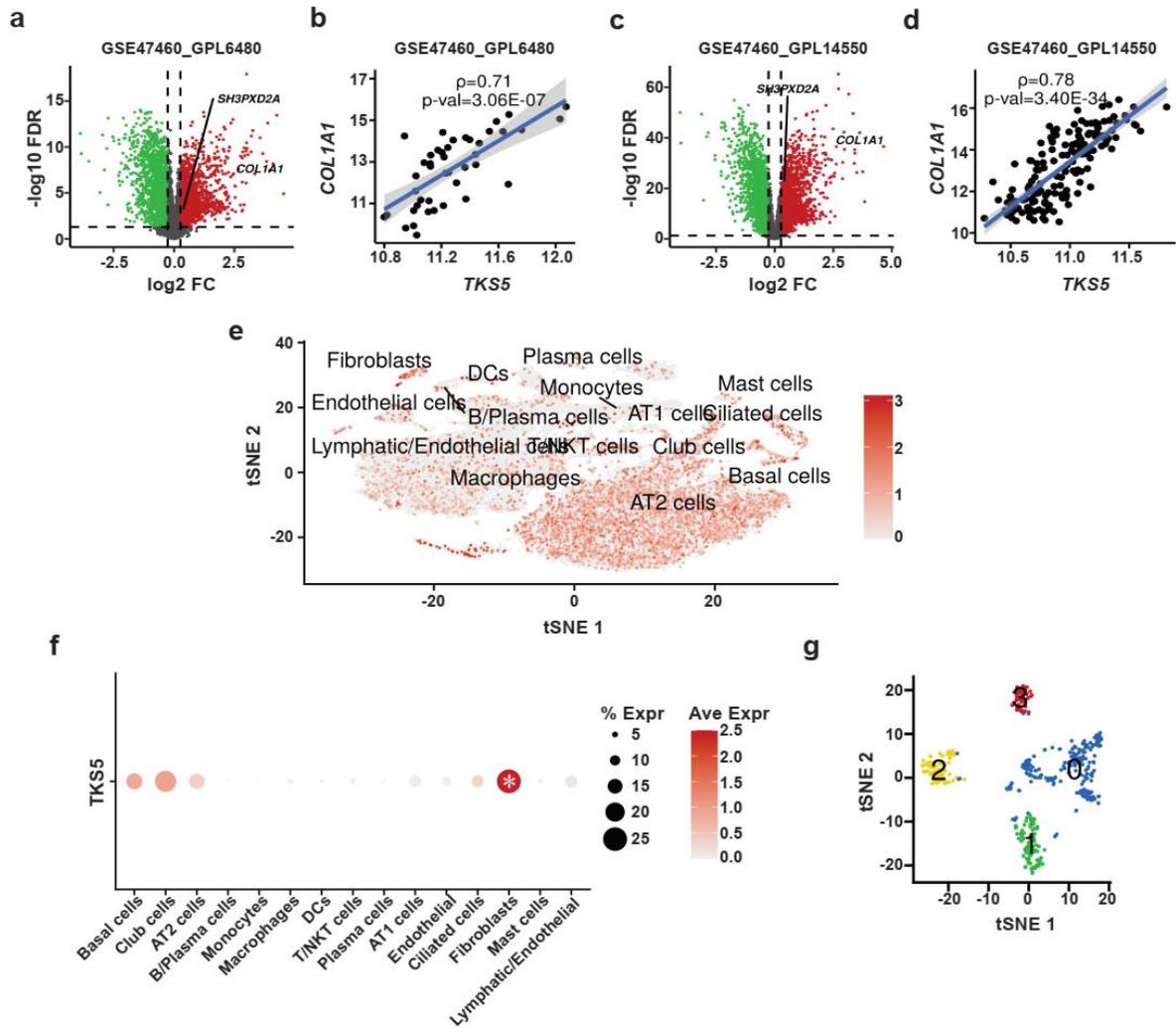


Figure S6. *TKS5* transcription is up-regulated in the lungs of IPF patients. **A, C.** DEA volcano plots between IPF and control lung tissue samples from two big publicly available transcriptomic datasets ($FC > 1.2$, $FDR < 0.05$). **B, D.** *TKS5* with *COL1A1* expression correlation in the A, C datasets. **E.** *TKS5* expression mapped in lung tissue cells embedded in a space of reduced dimensions; reanalysis of (Reyfan, Walter et al. 2019). **F.** Fibroblasts of the dataset in (E) are the primary producers of *TKS5* (Wilcoxon rank sum test; $FC > 1.2$, Bonferroni adjusted- $p < 0.05$). **G.** Same dataset fibroblast sub-groups with a resolution of 0.1 on isolated re-processed cells. Adopted from (Barbayianni, Kanellopoulou et al. 2023).

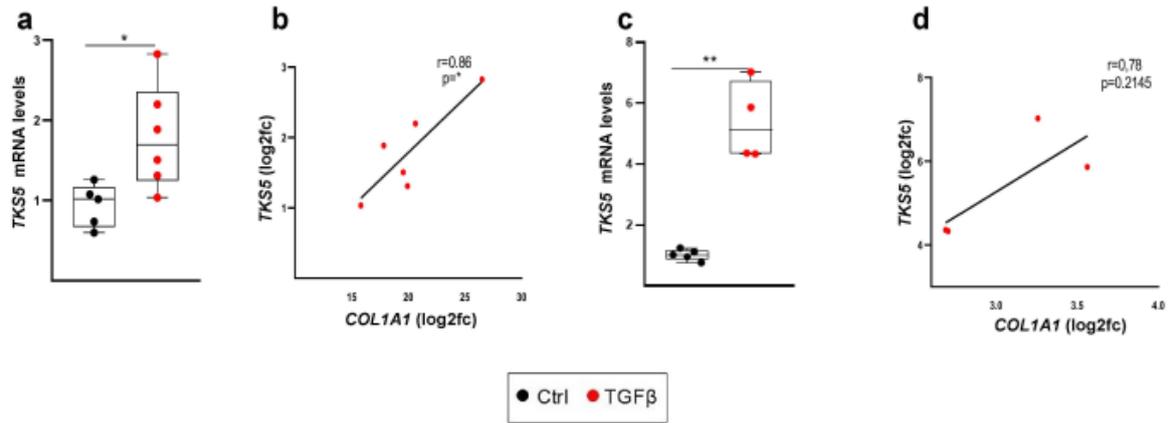


Figure S7. Lung fibroblasts over-express *TKS5* and *COL1A1* after stimulation with TGF β . **A.** Q-RT-PCR measured expression of *TKS5* in serum-depleted primary normal human lung fibroblasts cultured with TGF β . Expression values were normalized over those of *B2M* and presented as fold change over control (two-tailed t-test; * $p = 0.023$). **B.** *TKS5*-*COL1A1* expression correlation (two-tailed Pearson's correlation test; * $p = 0.0286$). **C.** Q-RT-PCR measured expression of *TKS5* in serum-depleted MRC5 cells (human lung fibroblasts) cultured with TGF β . Expression values were normalized over those of *B2M* and presented as fold change over control (two-tailed Welch's test; ** $p = 0.062$). **D.** *TKS5*-*COL1A1* expression correlation (two-tailed Pearson's correlation test). Adapted from (Barbayianni, Kanellopoulou et al. 2023).

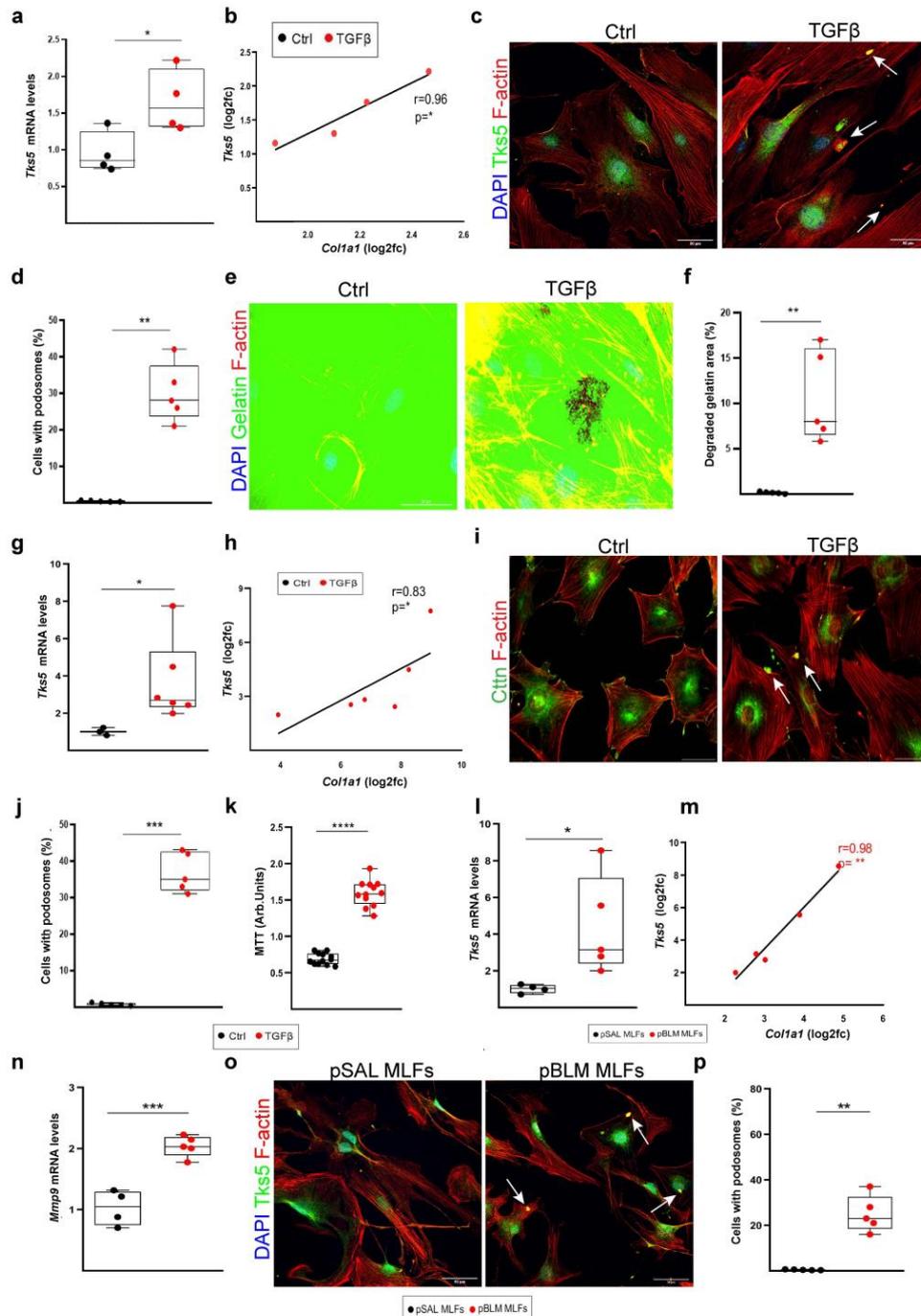


Figure S8. Podosome rosettes are inherently formed by BLM-treated lung fibroblasts post to TGFβ treatment. A-B. Q-RT-PCR measured *Tks5* and *Col1a1* expression by normal mouse lung fibroblasts with or without TGFβ treatment. The two genes follow the same transcription pattern post to TGFβ treatment. Expression was normalized over that of *B2m* and presented as fold change over control (two-tailed t-test with * $p=0.0322$; two-tailed Pearson's correlation test with * $p=0.0186$ and $r=0.96$). C. F-actin and *Tks5* double immunostaining images (scale bars 50μm; arrows indicate representative podosomes). D. Number of podosome-decorated cells per optical field ($n=5$; two-tailed Welch's test; ** $p=0.0012$). E. Images of normal mouse lung fibroblasts degrading a fluorescein-conjugated gelatin substrate (black holes) post to TGFβ treatment (scale bars 50μm). F. Quantification of gelatin degradation (ImageJ; two-tailed Welch's test; ** $p=0.0098$). G-H. *Tks5* and *Col1a1* expression from 3T3 cells with or without TGFβ treatment (same experimental procedure with A-B; two-tailed Mann-Whitney test with * $p=0.0238$; two-tailed Spearman's correlation with * $p=0.0167$ and

r=0.83). **I.** F-actin and cortactin (Ctn) double immunostaining images (scale bars 50 μ m; arrows indicate representative podosomes). **J.** Number of podosome-decorated cells per optical field (n=5; two-tailed Welch's test; ***p=0.0001). **K.** 3T3 cellular proliferation post to TGF β treatment (MTT assay; two-tailed Welch's test; ****p<0.0001). **L-M.** *Tks5* and *Colla1* expression from normal mouse fibroblasts post BLM administration (same experimental procedure with A-B; two-tailed Welch's test with *p=0.0464; two-tailed Pearson's correlation with **p=0.0023 and r=0.98). **N.** *Mmp9* expression measured as in L (two-tailed t-test, ***p=0.0003). **O.** F-actin and *Tks5* double immunostaining images (scale bars 50 μ m; arrows indicate representative podosomes). **P.** Number of podosome-decorated cells per optical field (two-tailed Welch's test; **p=0.0023). Adapted from (Barbayianni, Kanellopoulou et al. 2023).

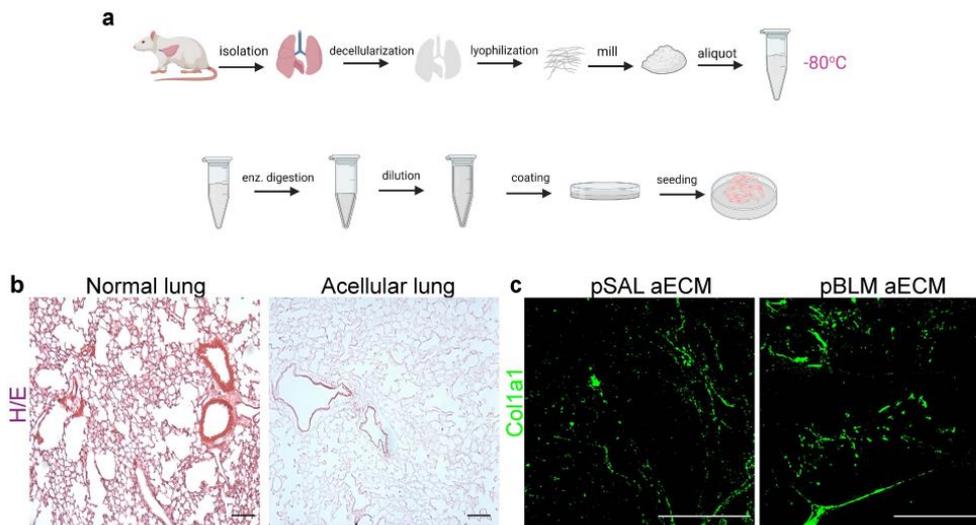


Figure S9. Usage of acellular ECM (aECM) as a substrate for autologous *in vitro* lung fibroblast culturing. **A.** Pipeline of aECM preparation from mouse lungs. **B.** H&E staining confirms decellularization of the collected matrix (scale bars 50 μ m). **C.** *Col1a1* immunofluorescent staining of aECM generated from mouse lungs treated with saline or BLM. The latter treatment induces the accumulation of collagen (scale bars 50 μ m). Adapted from (Barbayianni, Kanellopoulou et al. 2023).

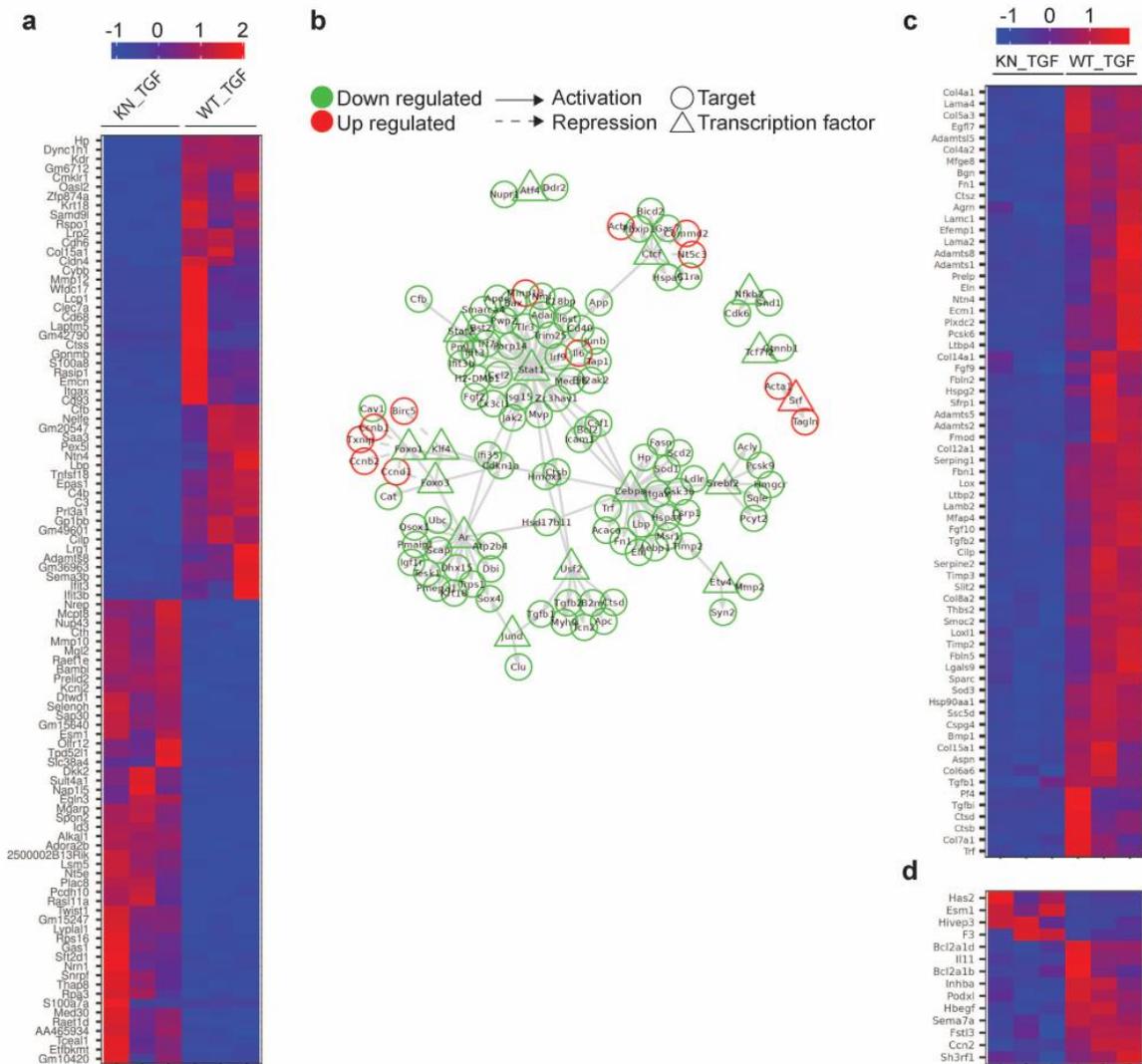
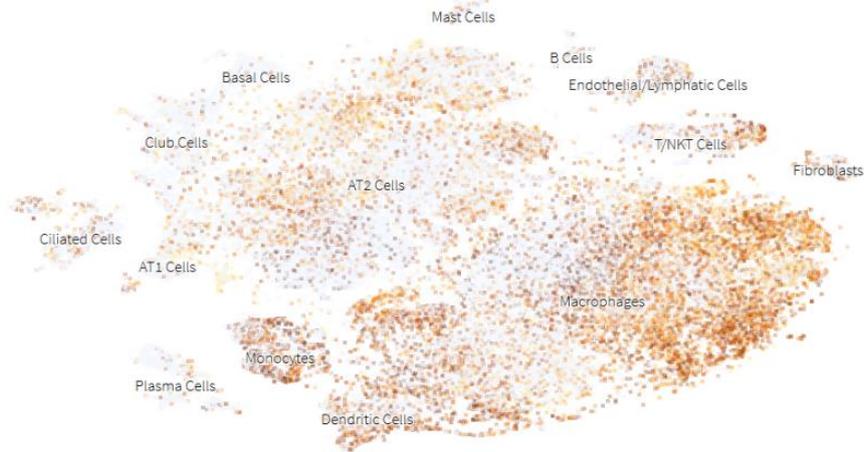
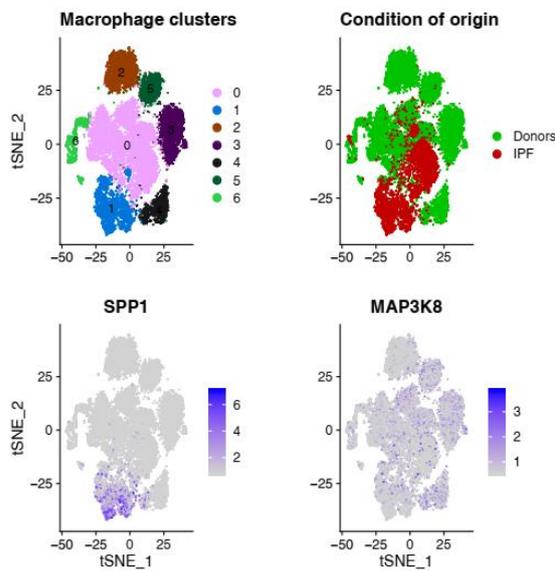


Figure S10. Expression patterns of *Tks5*^{+/-} lung fibroblasts treated with TGFβ. A. Top 100 DEGs (FC≥1.2; Benjamini-Hochberg corrected PANDORA p<0.05). **B.** Interaction of differentially expressed TFs and their target genes. **C.** Expression motif of core features from the GO:0062023 term selected post to pathway analysis. **D.** Deregulated mouse homologues of the top 50 up-regulated genes measured in invasive fibroblasts (supplementary data of PMID 30763282). Adopted from (Barbayaanni, Kanellopoulou et al. 2023).

A



B



C

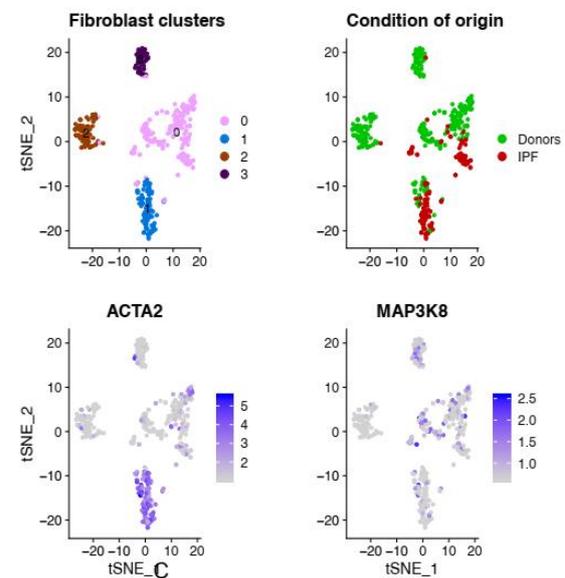


Figure S11. Cell specificity of *MAP3K8* expression in the human lung. The gene is down-regulated in IPF macrophages and myofibroblasts. **A.** Projection of 76,070 single cells in reduced dimension space. Cells were isolated from eight donor lung biopsies and equal number of IPF patients' lung explants. 14 different cell types were identified (nupulmonary.org/resources; GSE122960). **B-C.** Re-analysis of the same dataset. **B.** *MAP3K8* is under-expressed in macrophage sub-cluster 1 that is characterized by *SPP1* expression and cells of mainly IPF origin (FC = -1.19; Bonferroni adjusted p-value = 6.11e-24). **C.** Fibroblast sub-clustering from the same dataset. *MAP3K8* is under-expressed in aSMA+ sub-cluster 1 (FC=-1.35; Bonferroni adjusted p-value = 0.0002). Adopted from (Zannikou, Barbayanni et al. 2021).

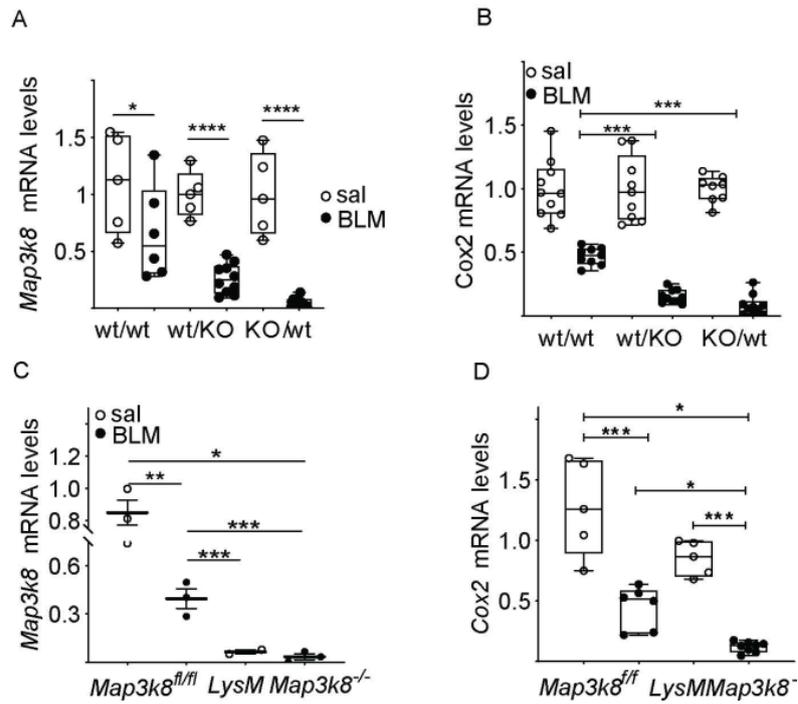


Figure S12. *Cox-2* expression in the lungs is regulated by *Map3k8*. **A, C.** Q-RT-PCR measured *Map3k8* and **(B, D)** *Cox-2* expression in whole lung tissue of the indicated conditions, normalized over the expression of *B2m* (two-way ANOVA with Bonferroni post-hoc correction; */**/**** p = 0.05/0.01/0.001, respectively). Adopted from (Zannikou, Barbayianni et al. 2021).

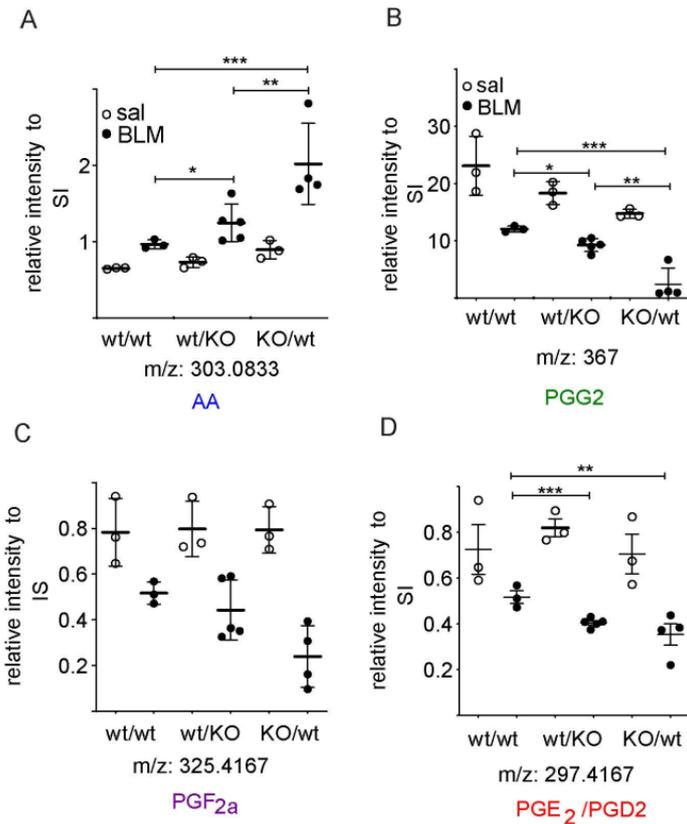


Figure S13. Map3k8 is a regulator of arachidonic acid and PGE₂. Arachidonic acid (A) and PGE₂ (B) levels in the BAL fluid of chimeric mice as determined by MS/MS (two-way ANOVA with Bonferroni's post-hoc correction; */**/** p = 0.05/0.01/0.001, respectively). Adopted from (Zannikou, Barbayianni et al. 2021).

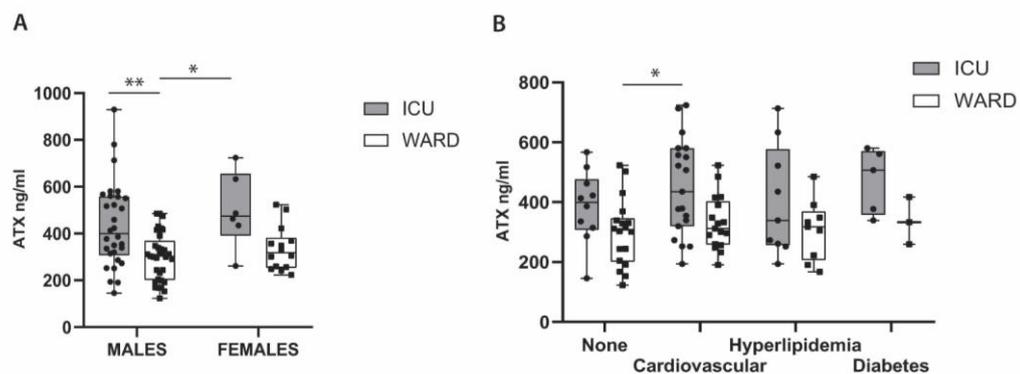


Figure S14. ATX levels in the serum of ICU COVID-19 patients do not correlate with their gender or underlying comorbidities. Serum ATX was measured using ELISA in the sera of ward and ICU patients of the Evaggelismos hospital. Patients were treated or not with Dex. Gender (A) or existing comorbidities did not seem to affect ATX levels (two-way ANOVA with Bonferroni's post-hoc correction; */** p < 0.001/0.0001, respectively). Adopted from (Nikitopoulou, Fanidis et al. 2021).

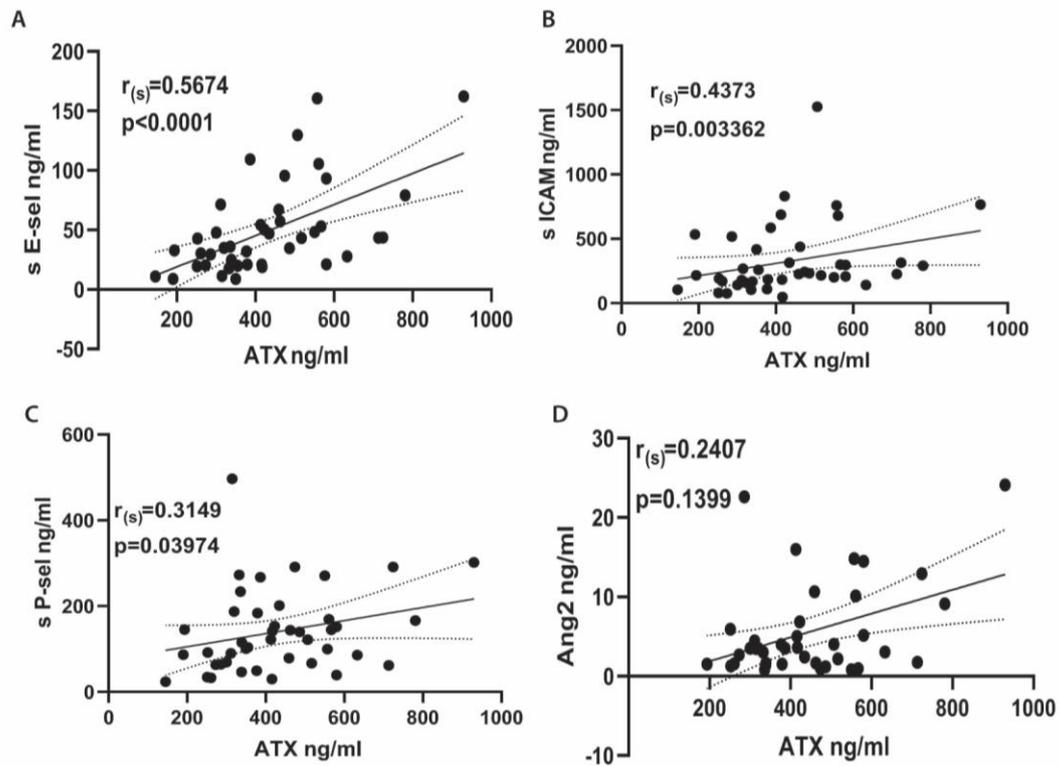


Figure S15. Levels of endothelial dysfunction markers positively correlate with ATX levels in the serum of ICU-admitted COVID-19 patients. E-sel (A), sICAM (B), sP-sel (C) and ANG2 (D) levels positively correlate with ATX protein levels in the serum of SARS-CoV-2 affected individuals (Spearman's correlation). Adopted from (Nikitopoulou, Fanidis et al. 2021).

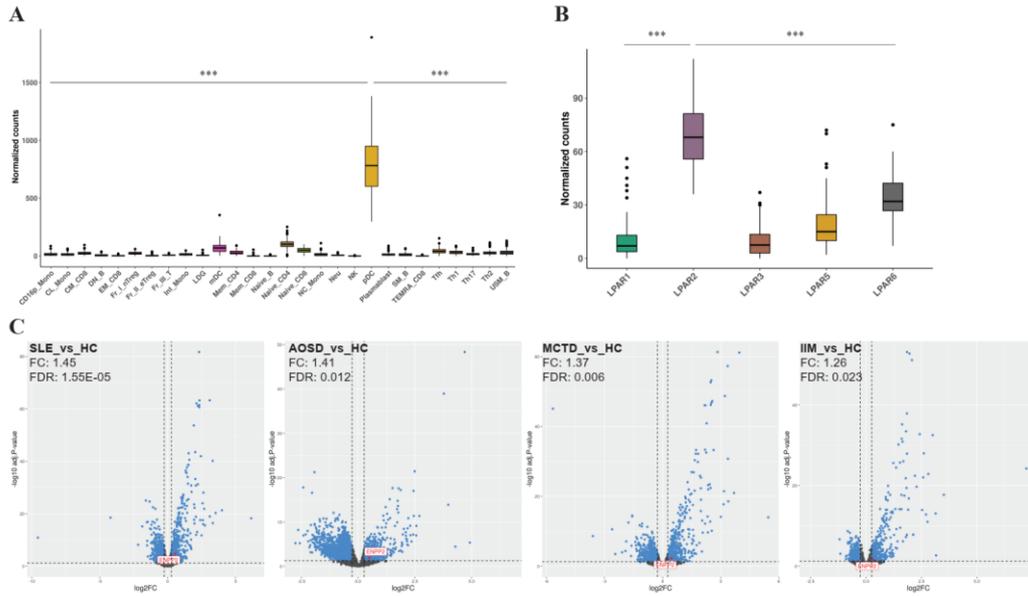


Figure S17. *ENPP2* expression by plasmacytoid dendritic cells (pDCs) is deregulated during inflammatory diseases. **A.** pDCs are the main *ENPP2* producers among immune cells during steady state. Per cell type EDASeq-normalized counts and pairwise differential expression analysis of each cell type with pDCs (***) iconizes $FC > 1.2$ and FDR-corrected $p < 0.01$. **B.** *LPAR2* is the main LPA receptor of healthy pDCs (EDASeq-normalized counts; Kruskal-Wallis followed by Dunn post-hoc test; *** denotes Benjamini-Hochberg adjusted $p < 0.01$). **C.** pDCs sampled from four immune-related pathologies have a more intense *ENPP2* transcription compared to their healthy counterparts ($FC > 1.2$; FDR-corrected $p < 0.05$; PMID: 33930287). HC: healthy control; SLE: systemic lupus erythematosus; AOSD: adult-onset Still's disease; MCTD: mixed connective tissue disease; IIM: idiopathic inflammatory myopathy. Adopted from (Nikitopoulou, Fanidis et al. 2021).

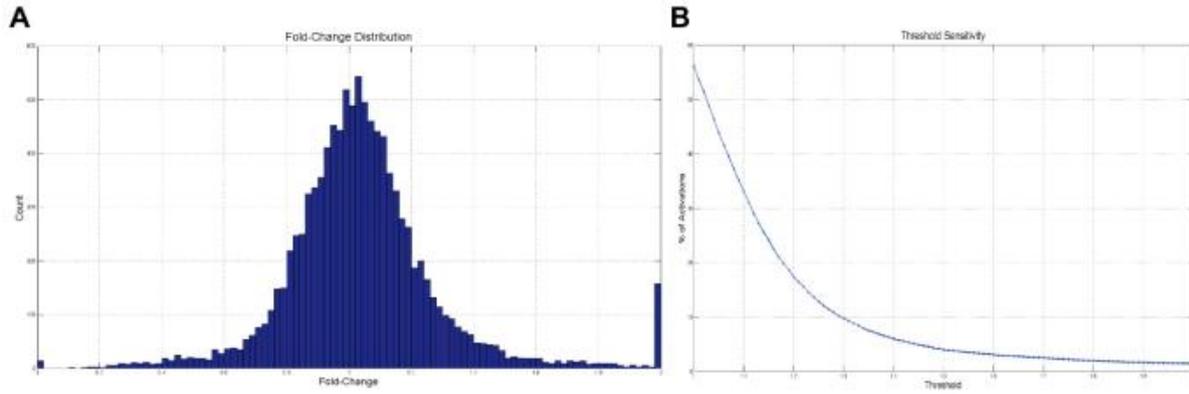


Figure S18. Selection of an active signal-response threshold. **A.** Fold change (measured fluorescence under stimulated to unstimulated state) distribution for all signals. Median response value post to stimulation with culture medium was selected as the unstimulated state. **B.** Fold change threshold sensitivity analysis. Selection of a small threshold has a big impact on the number of active signals filtered in, whereas high thresholds render data rather insensitive to threshold changes. Optimal threshold was selected as the one which enabled an adequate number of signals for downstream analysis but at the same time made the dataset insensitive enough to ratio changes. Adopted from (Magkrioti, Antonopoulou et al. 2022).

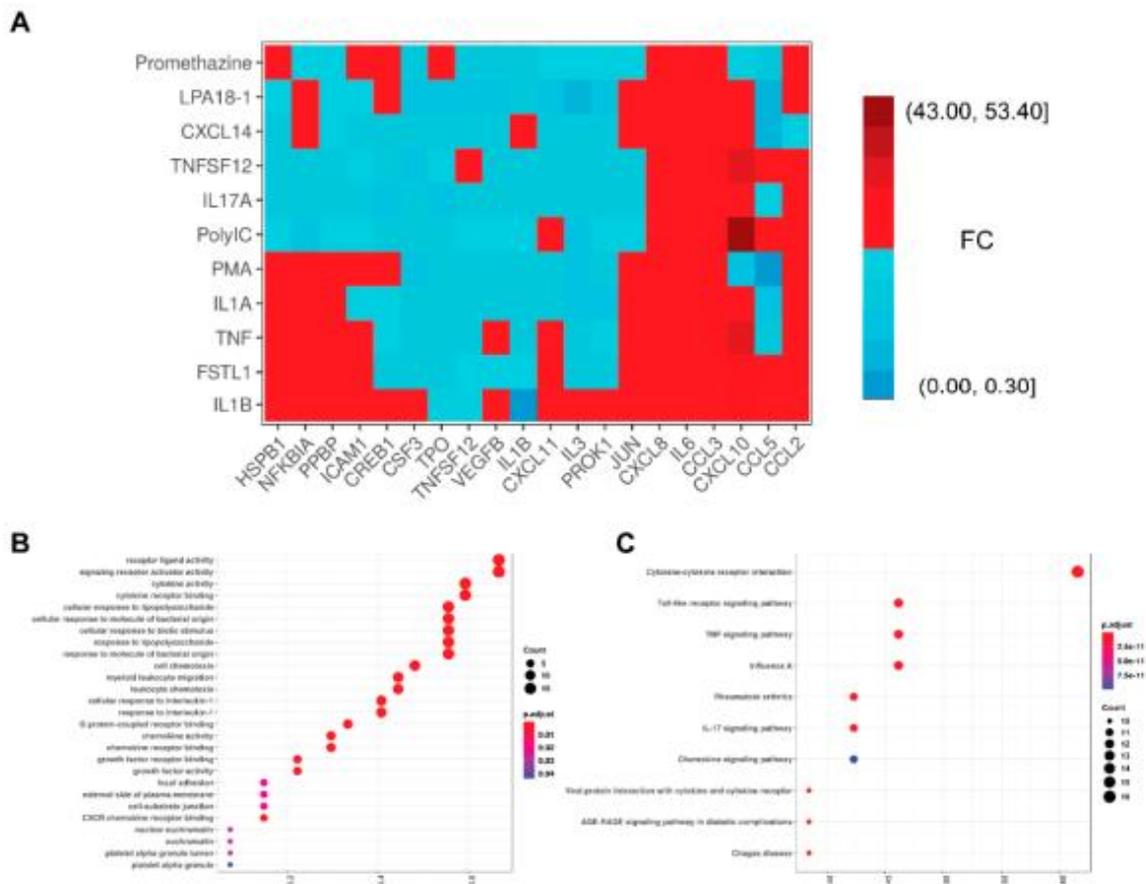


Figure S19. LPA impacts renal proximal tubular epithelial cells similarly to pro-inflammatory stimuli. **A.** Heatmap of cluster II (totally unresponsive signals were filtered out). **B-C.** Pathway analysis of cluster II signals/stimuli with GO and KEGG reference pathways, respectively. Top 10 enriched

terms from KEGG and each GO category reveal the pro-inflammatory nature of the cluster, suggesting LPA pro-inflammatory *modus operandi*. Adopted from (Magkrioti, Antonopoulou et al. 2022).

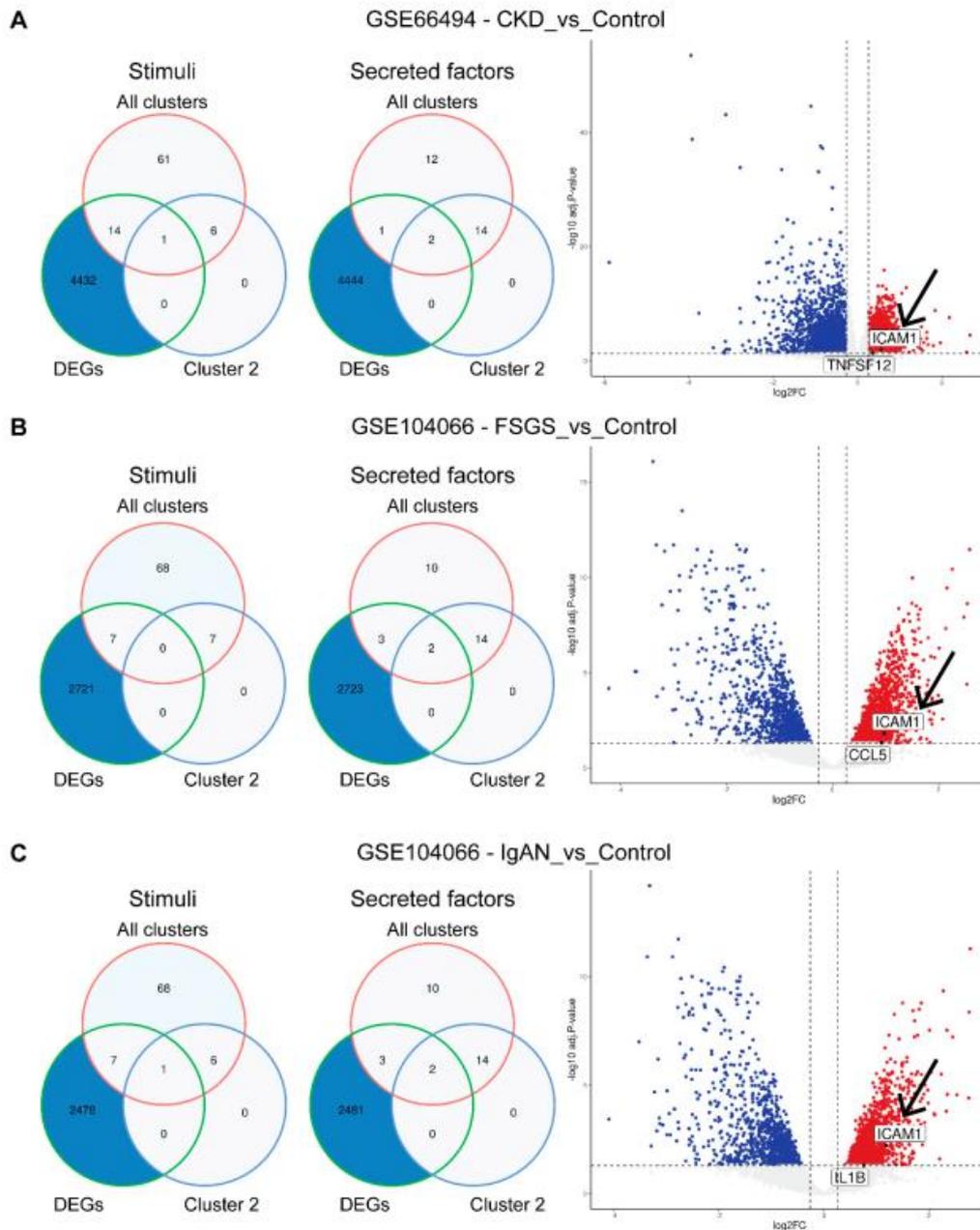


Figure S20. Transcription profile of cluster II targets upon kidney disease. Re-analysis of a couple of microarray datasets (GSE66494 in A and GSE104066 in B-C) representing steady state and three kidney pathologies (CKD: chronic kidney disease; FGS: focal segmental glomerulosclerosis; IgAN: immunoglobulin A nephropathy). Venn diagrams iconize the similarities of differentially expressed genes ($|FC| \geq 1.2$ with adjusted p -value < 0.05) between shown comparisons with the ELISA-measured active stimuli and secreted molecules. Intersection features are shown on the volcano plots with arrows indicating the LPA responders. Interestingly, TNFSF12 and IL1B1 are both stimuli and responders of cluster II. Adopted from (Magkrioti, Antonopoulou et al. 2022).

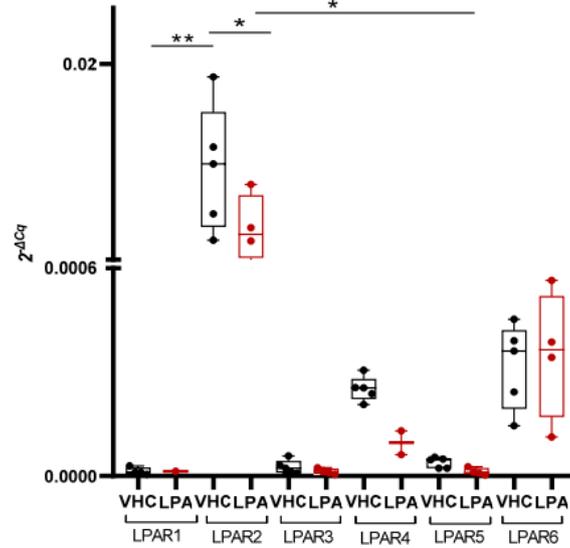


Figure S22. Endogenous levels of LPA receptors (LPARs) transcription from HKC-8 cells. Cells were treated with LPA 18:1 or chloroform (VHC). Expression levels were assessed with RT-qPCR and were normalized over those of *B2M* prior to comparison with $2^{-\Delta Cq}$ values (Kruskal-Wallis test; $^{**}/^{*}p < 0.05/0.01$). Adopted from (Magkrioti, Antonopoulou et al. 2022).

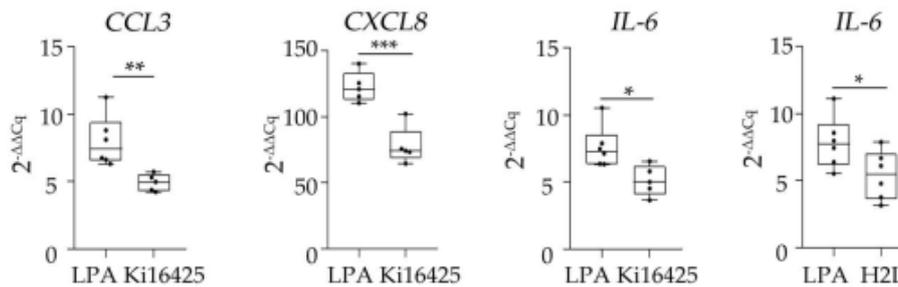


Figure S23. LPA receptors (LPARs) as mediators of LPA signaling. HKC-8 cells were treated with $10\mu\text{M}$ of Ki16425 (LPAR1/3 inhibitor) or equal amount of H2L5186303 (LPAR2 inhibitor) post to LPA treatment at a final concentration of $10\mu\text{M}$ for 4h. Unpaired t-test or Welch's test were used in case of normally distributed data, while Mann-Whitney otherwise; $^{**}/^{**}/^{***}p < 0.05/0.01/0.001$. Adopted from (Magkrioti, Antonopoulou et al. 2022).

Appendix II – Supplementary tables

Table S1. Fibromine-hosted bulk sequencing datasets.

Datasets	
Transcriptomics	51
Human	38
Mouse	13
Proteomics	9
Human	9
Mouse	-
Publications	
Transcriptomics	47
Proteomics	9
Comparisons	
Transcriptomics	42
Proteomics	1
Tissue/cell types	
Transcriptomics	15
Proteomics	4

Table S2. Differentially expressed features used for machine learning training purposes. Adapted from (Fanidis, Pezoulas et al. 2023).

Symbol	Description
<i>ABCA3</i>	ATP binding cassette subfamily A member 3
<i>ABHD5</i>	abhydrolase domain containing 5, lysophosphatidic acid acyltransferase
<i>ACADL</i>	acyl-CoA dehydrogenase long chain
<i>ACSL1</i>	acyl-CoA synthetase long chain family member 1
<i>ACVRL1</i>	activin A receptor like type 1
<i>ADAMDEC1</i>	ADAM like decysin 1
<i>AGER</i>	advanced glycosylation end-product specific receptor
<i>AGTR2</i>	angiotensin II receptor type 2
<i>AKR1A1</i>	aldo-keto reductase family 1 member A1
<i>ALPP</i>	alkaline phosphatase, placental
<i>ANKRD1</i>	ankyrin repeat domain 1
<i>ANXA3</i>	annexin A3
<i>AQP5</i>	aquaporin 5
<i>ARG1</i>	arginase 1
<i>ARNTL2</i>	aryl hydrocarbon receptor nuclear translocator like 2
<i>ASPN</i>	asporin
<i>B3GNT3</i>	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 3
<i>BDKRB2</i>	bradykinin receptor B2
<i>BHLHE22</i>	basic helix-loop-helix family member e22
<i>BPIFB1</i>	BPI fold containing family B member 1
<i>C20orf85</i>	chromosome 20 open reading frame 85
<i>CA2</i>	carbonic anhydrase 2
<i>CA4</i>	carbonic anhydrase 4
<i>CACNA2D2</i>	calcium voltage-gated channel auxiliary subunit alpha2delta 2
<i>CACNB4</i>	calcium voltage-gated channel auxiliary subunit beta 4
<i>CALU</i>	calumenin

<i>CCL11</i>	C-C motif chemokine ligand 11
<i>CCL13</i>	C-C motif chemokine ligand 13
<i>CCL19</i>	C-C motif chemokine ligand 19
<i>CD19</i>	CD19 molecule
<i>CD1A</i>	CD1a molecule
<i>CD1C</i>	CD1c molecule
<i>CD1E</i>	CD1e molecule
<i>CHST6</i>	carbohydrate sulfotransferase 6
<i>CLDN18</i>	claudin 18
<i>COL15A1</i>	collagen type XV alpha 1 chain
<i>COL17A1</i>	collagen type XVII alpha 1 chain
<i>COL1A1</i>	collagen type I alpha 1 chain
<i>COL1A2</i>	collagen type I alpha 2 chain
<i>COL3A1</i>	collagen type III alpha 1 chain
<i>COL5A2</i>	collagen type V alpha 2 chain
<i>COLEC11</i>	collectin subfamily member 11
<i>COMP</i>	cartilage oligomeric matrix protein
<i>COX4I2</i>	cytochrome c oxidase subunit 4I2
<i>CP</i>	ceruloplasmin
<i>CPB2</i>	carboxypeptidase B2
<i>CRLF1</i>	cytokine receptor like factor 1
<i>CSF3</i>	colony stimulating factor 3
<i>CSF3R</i>	colony stimulating factor 3 receptor
<i>CTH</i>	cystathionine gamma-lyase
<i>CTHRC1</i>	collagen triple helix repeat containing 1
<i>CTSG</i>	cathepsin G
<i>CXCL14</i>	C-X-C motif chemokine ligand 14
<i>CXCL6</i>	C-X-C motif chemokine ligand 6
<i>CXCR1</i>	C-X-C motif chemokine receptor 1
<i>CYBRD1</i>	cytochrome b reductase 1
<i>CYP24A1</i>	cytochrome P450 family 24 subfamily A member 1
<i>CYP2F1</i>	cytochrome P450 family 2 subfamily F member 1
<i>DAPK2</i>	death associated protein kinase 2
<i>DCXR</i>	dicarbonyl and L-xylulose reductase
<i>DISP1</i>	dispatched RND transporter family member 1
<i>DNAAF1</i>	dynein axonemal assembly factor 1
<i>DPYSL3</i>	dihydropyrimidinase like 3
<i>DYNLRB2</i>	dynein light chain roadblock-type 2
<i>EMP2</i>	epithelial membrane protein 2
<i>F11</i>	coagulation factor XI
<i>FERMT2</i>	fermitin family member 2
<i>FLVCR2</i>	FLVCR heme transporter 2
<i>FOXF1</i>	forkhead box F1
<i>FOXJ1</i>	forkhead box J1
<i>GDF15</i>	growth differentiation factor 15
<i>GKN2</i>	gastrokine 2

<i>GPD1L</i>	glycerol-3-phosphate dehydrogenase 1 like
<i>GPM6B</i>	glycoprotein M6B
<i>GPR87</i>	G protein-coupled receptor 87
<i>GPX3</i>	glutathione peroxidase 3
<i>HDC</i>	histidine decarboxylase
<i>HEY1</i>	hes related family bHLH transcription factor with YRPW motif 1
<i>HIF3A</i>	hypoxia inducible factor 3 subunit alpha
<i>HIVEP3</i>	HIVEP zinc finger 3
<i>HS6ST2</i>	heparan sulfate 6-O-sulfotransferase 2
<i>HSD17B6</i>	hydroxysteroid 17-beta dehydrogenase 6
<i>ICAM2</i>	intercellular adhesion molecule 2
<i>IL13RA2</i>	interleukin 13 receptor subunit alpha 2
<i>IL18R1</i>	interleukin 18 receptor 1
<i>IL18RAP</i>	interleukin 18 receptor accessory protein
<i>IL1R2</i>	interleukin 1 receptor type 2
<i>IL1RL1</i>	interleukin 1 receptor like 1
<i>INPP1</i>	inositol polyphosphate-1-phosphatase
<i>ITGB8</i>	integrin subunit beta 8
<i>KCNK3</i>	potassium two pore domain channel subfamily K member 3
<i>KLF9</i>	Kruppel like factor 9
<i>KRT15</i>	keratin 15
<i>LCN2</i>	lipocalin 2
<i>LHPP</i>	phospholysine phosphohistidine inorganic pyrophosphate phosphatase
<i>LSS</i>	lanosterol synthase
<i>MAP3K6</i>	mitogen-activated protein kinase kinase kinase 6
<i>MFSD2A</i>	major facilitator superfamily domain containing 2A
<i>MGAM</i>	maltase-glucoamylase
<i>MLPH</i>	melanophilin
<i>MMP1</i>	matrix metalloproteinase 1
<i>MMP7</i>	matrix metalloproteinase 7
<i>MS4A2</i>	membrane spanning 4-domains A2
<i>MSMB</i>	microseminoprotein beta
<i>MSMO1</i>	methylsterol monooxygenase 1
<i>MVD</i>	mevalonate diphosphate decarboxylase
<i>MYLK</i>	myosin light chain kinase
<i>MZB1</i>	marginal zone B and B1 cell specific protein
<i>NDRG4</i>	NDRG family member 4
<i>NEK11</i>	NIMA related kinase 11
<i>NFATC4</i>	nuclear factor of activated T cells 4
<i>NTF3</i>	neurotrophin 3
<i>PAPSS2</i>	3'-phosphoadenosine 5'-phosphosulfate synthase 2
<i>PARD3</i>	par-3 family cell polarity regulator
<i>PAX9</i>	paired box 9
<i>PCTP</i>	phosphatidylcholine transfer protein
<i>PDLIM4</i>	PDZ and LIM domain 4
<i>PEBP4</i>	phosphatidylethanolamine binding protein 4

<i>PGC</i>	progastricin
<i>PIGA</i>	phosphatidylinositol glycan anchor biosynthesis class A
<i>PLA2G1B</i>	phospholipase A2 group IB
<i>PLA2G2A</i>	phospholipase A2 group IIA
<i>PLA2G7</i>	phospholipase A2 group VII
<i>PROK2</i>	prokineticin 2
<i>PRX</i>	periaxin
<i>PTPRB</i>	protein tyrosine phosphatase receptor type B
<i>PTPRE</i>	protein tyrosine phosphatase receptor type E
<i>RAMP3</i>	receptor activity modifying protein 3
<i>RCAN2</i>	regulator of calcineurin 2
<i>RENBP</i>	renin binding protein
<i>RGMA</i>	repulsive guidance molecule BMP co-receptor a
<i>RGS10</i>	regulator of G protein signaling 10
<i>RGS16</i>	regulator of G protein signaling 16
<i>RNF128</i>	ring finger protein 128
<i>S100A2</i>	S100 calcium binding protein A2
<i>S100A8</i>	S100 calcium binding protein A8
<i>SCEL</i>	sciellin
<i>SCGB1A1</i>	secretoglobin family 1A member 1
<i>SCN7A</i>	sodium voltage-gated channel alpha subunit 7
<i>SERPIND1</i>	serpin family D member 1
<i>SFRP2</i>	secreted frizzled related protein 2
<i>SFTPD</i>	surfactant protein D
<i>SIX1</i>	SIX homeobox 1
<i>SLAMF7</i>	SLAM family member 7
<i>SLC19A2</i>	solute carrier family 19 member 2
<i>SLC6A4</i>	solute carrier family 6 member 4
<i>SLCO1A2</i>	solute carrier organic anion transporter family member 1A2
<i>SLCO2A1</i>	solute carrier organic anion transporter family member 2A1
<i>SLN</i>	sarcolipin
<i>SOSTDC1</i>	sclerostin domain containing 1
<i>SOX7</i>	SRY-box transcription factor 7
<i>SPP1</i>	secreted phosphoprotein 1
<i>SPRY4</i>	sprouty RTK signaling antagonist 4
<i>SRXN1</i>	sulfiredoxin 1
<i>ST6GALNAC1</i>	ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 1
<i>STEAP1</i>	STEAP family member 1
<i>STRA6</i>	signaling receptor and transporter of retinol STRA6
<i>SULF1</i>	sulfatase 1
<i>TACC2</i>	transforming acidic coiled-coil containing protein 2
<i>THBS4</i>	thrombospondin 4
<i>TMPRSS4</i>	transmembrane serine protease 4
<i>TNFRSF17</i>	TNF receptor superfamily member 17
<i>TNNC1</i>	troponin C1, slow skeletal and cardiac type
<i>TRO</i>	trophinin

<i>UNC13B</i>	unc-13 homolog B
<i>VCAM1</i>	vascular cell adhesion molecule 1
<i>VIPR1</i>	vasoactive intestinal peptide receptor 1
<i>VNN2</i>	vanin 2
<i>VWCE</i>	von Willebrand factor C and EGF domains
<i>WFDC2</i>	WAP four-disulfide core domain 2
<i>XYLT2</i>	xylosyltransferase 2
<i>ZBTB16</i>	zinc finger and BTB domain containing 16

Table S3. GO terms enriched in the 76 features prioritized by machine learning. Adapted from (Fanidis, Pezoulas et al. 2023).

ID	Description
GO:0030020	extracellular matrix structural constituent conferring tensile strength
GO:0005201	extracellular matrix structural constituent
GO:0005581	collagen trimer
GO:0062023	collagen-containing extracellular matrix
GO:0005583	fibrillar collagen trimer
GO:0098643	banded collagen fibril
GO:0005788	endoplasmic reticulum lumen
GO:0048407	platelet-derived growth factor binding
GO:0048771	tissue remodeling
GO:0019955	cytokine binding
GO:0098644	complex of collagen trimers
GO:0005178	integrin binding
GO:0030199	collagen fibril organization
GO:0043062	extracellular structure organization
GO:0030198	extracellular matrix organization
GO:0008009	chemokine activity
GO:0071559	response to transforming growth factor beta
GO:0071560	cellular response to transforming growth factor beta stimulus
GO:0042476	odontogenesis
GO:0019730	antimicrobial humoral response
GO:0042379	chemokine receptor binding
GO:0002020	protease binding
GO:0019838	growth factor binding
GO:0046332	SMAD binding
GO:0005125	cytokine activity
GO:0030593	neutrophil chemotaxis
GO:0004896	cytokine receptor activity
GO:0071723	lipopeptide binding
GO:0071347	cellular response to interleukin-1
GO:0032963	collagen metabolic process
GO:0048020	CCR chemokine receptor binding
GO:0048247	lymphocyte chemotaxis
GO:1990266	neutrophil migration
GO:0009897	external side of plasma membrane
GO:0071621	granulocyte chemotaxis
GO:0002709	regulation of T cell mediated immunity
GO:0042493	response to drug
GO:0071356	cellular response to tumor necrosis factor
GO:0070555	response to interleukin-1
GO:0001664	G protein-coupled receptor binding
GO:0001910	regulation of leukocyte mediated cytotoxicity
GO:0097529	myeloid leukocyte migration
GO:0010876	lipid localization

GO:0001914 regulation of T cell mediated cytotoxicity
 GO:0034612 response to tumor necrosis factor
 GO:1904018 positive regulation of vasculature development
 GO:0030595 leukocyte chemotaxis
 GO:0002822 regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains
 GO:0097530 granulocyte migration
 GO:0031341 regulation of cell killing
 GO:0034103 regulation of tissue remodeling
 GO:0046849 bone remodeling
 GO:0043506 regulation of JUN kinase activity
 GO:0048545 response to steroid hormone
 GO:0009636 response to toxic substance
 GO:0031960 response to corticosteroid
 GO:0031214 biomineral tissue development
 GO:0110148 biomineralization
 GO:0004089 carbonate dehydratase activity
 GO:0071674 mononuclear cell migration
 GO:0031667 response to nutrient levels
 GO:0007178 transmembrane receptor protein serine/threonine kinase signaling pathway
 GO:0007584 response to nutrient
 GO:0002819 regulation of adaptive immune response
 GO:0001906 cell killing
 GO:0045124 regulation of bone resorption
 GO:0015718 monocarboxylic acid transport
 GO:0001913 T cell mediated cytotoxicity
 GO:0002460 adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains
 GO:0015908 fatty acid transport
 GO:0006959 humoral immune response
 GO:0002456 T cell mediated immunity
 GO:0033197 response to vitamin E
 GO:0043589 skin morphogenesis
 GO:0016054 organic acid catabolic process
 GO:0046395 carboxylic acid catabolic process
 GO:0001909 leukocyte mediated cytotoxicity
 GO:0046850 regulation of bone remodeling
 GO:0006654 phosphatidic acid biosynthetic process
 GO:0006631 fatty acid metabolic process
 GO:0006869 lipid transport
 GO:0044272 sulfur compound biosynthetic process
 GO:0046322 negative regulation of fatty acid oxidation
 GO:0046473 phosphatidic acid metabolic process
 GO:0001912 positive regulation of leukocyte mediated cytotoxicity
 GO:0051216 cartilage development
 GO:0019755 one-carbon compound transport
 GO:0043200 response to amino acid

GO:0140375	immune receptor activity
GO:0072676	lymphocyte migration
GO:0001503	ossification
GO:0007179	transforming growth factor beta receptor signaling pathway
GO:0060348	bone development
GO:0002475	antigen processing and presentation via MHC class Ib
GO:0038166	angiotensin-activated signaling pathway
GO:0045766	positive regulation of angiogenesis
GO:0060326	cell chemotaxis
GO:0002703	regulation of leukocyte mediated immunity
GO:0090177	establishment of planar polarity involved in neural tube closure
GO:0019216	regulation of lipid metabolic process
GO:0031343	positive regulation of cell killing
GO:1901342	regulation of vasculature development
GO:0071496	cellular response to external stimulus
GO:0002699	positive regulation of immune effector process
GO:0001101	response to acid chemical
GO:0042249	establishment of planar polarity of embryonic epithelium
GO:0045453	bone resorption
GO:0060135	maternal process involved in female pregnancy

Table S4. MSigDB hallmark set enriched in the 76 features prioritized by machine learning.
Adapted from (Fanidis, Pezoulas et al. 2023).

ID	Description
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION

Table S5. Protein complexes of the 76 machine learning prioritized features potentially affected by IPF changes (CORUM3.0). Adapted from (Fanidis, Pezoulas et al. 2023).

Complex Name	Prioritized genes frequency	Prioritized gene(s) in complex
ITGAV-ITGB5-SPP1 complex	0.33	SPP1
ITGAV-ITGB6-SPP1 complex	0.33	SPP1
ITGAV-ITGB3-SPP1 complex	0.33	SPP1
ITGAV-ITGB3-SPP1 complex	0.33	SPP1
ITGAV-ITGB1-SPP1 complex	0.33	SPP1
ITGA11-ITGB1-COL1A1 complex	0.33	COL1A1
Sulphiredoxin-peroxiredoxin complex	0.50	SRXN1
CR2-CD19 complex	0.50	CD19
CP-LF complex	0.50	CP
CP-MPO complex	0.50	CP
ITGAV-ITGB8-MMP14-TGFB1 complex	0.25	ITGB8
CD19-VAV1-PIK3R1 complex	0.33	CD19
ITGA9-ITGB1-SPP1 complex	0.33	SPP1
ITGA5-ITGB1-SPP1 complex	0.33	SPP1
Chromatin remodeling complex (TACC2, TACC3, PCAF)	0.33	TACC2
FBXW7-CSF3R-GSK3B complex	0.33	CSF3R
CP-LF-MPO complex	0.33	CP
CTNNB1-FERMT2-TCF7L2 complex	0.33	FERMT2
CTNNB1-FERMT2-YBX1 complex	0.33	FERMT2
HIVEP3-RUNX2-WWP1 complex	0.33	HIVEP3
P2X7 receptor signaling complex	0.08	PTPRB
Itga5-Itgb1-Fn1-Sfrp2 complex	0.25	SFRP2
MLL1-WDR5 complex	0.04	MGAM

Table S6. Relationship of lite model-used features to spirometry measurements. Table shows genes with significant relationships of the same direction in at least two datasets. Adapted from (Fanidis, Pezoulas et al. 2023).

Gene	Dataset	Spirometry measurement	Spearman rho	p value
ABCA3	GSE47460 GPL6480	% predicted DLCO	0.55	4.61E-04
		% predicted FEV1 post-bd	0.67	6.07E-03
	% predicted FEV1 pre-bd	0.36	1.90E-02	
	% predicted FVC post-bd	0.72	2.51E-03	
	% predicted FVC pre-bd	0.44	3.51E-03	
	(FEV1 / FVC) post-bd	-0.54	3.86E-02	
	(FEV1 / FVC) pre-bd	-0.41	6.40E-03	
	GSE47460 GPL14550	% predicted DLCO	0.50	1.74E-10
		% predicted FEV1 post-bd	0.29	6.10E-03
		% predicted FEV1 pre-bd	0.43	1.45E-08
		% predicted FVC post-bd	0.41	8.68E-05
		% predicted FVC pre-bd	0.54	3.14E-13
		(FEV1 / FVC) post-bd	-0.44	1.78E-05
	(FEV1 / FVC) pre-bd	-0.50	1.37E-11	
ABHD5	GSE47460 GPL6480	% predicted DLCO	0.42	1.06E-02
		% predicted DLCO	0.37	3.98E-06
	% predicted FVC pre-bd	0.40	1.86E-07	
	GSE32537	% predicted DLCO	-0.25	1.47E-02
		% predicted FVC pre-bd	-0.24	1.15E-02
ACADL	GSE47460 GPL6480	% predicted DLCO	0.65	1.40E-05
		% predicted FEV1 post-bd	0.77	8.15E-04
		% predicted FEV1 pre-bd	0.52	3.17E-04
		% predicted FVC post-bd	0.72	2.51E-03
		% predicted FVC pre-bd	0.60	2.24E-05
		(FEV1 / FVC) pre-bd	-0.48	1.03E-03
	GSE47460 GPL14550	% predicted DLCO	0.65	1.76E-18
		% predicted FEV1 post-bd	0.50	9.12E-07
		% predicted FEV1 pre-bd	0.57	3.03E-15
		% predicted FVC post-bd	0.62	1.50E-10
		% predicted FVC pre-bd	0.67	5.31E-22
		(FEV1 / FVC) pre-bd	-0.53	6.56E-13
		ADAMDEC1	GSE47460 GPL6480	% predicted FEV1 pre-bd
% predicted FVC pre-bd	-0.34			2.65E-02
GSE47460 GPL14550	% predicted FEV1 pre-bd		-0.31	5.61E-05
	% predicted FVC pre-bd		-0.37	1.90E-06
AGER	GSE47460 GPL6480	% predicted DLCO	0.74	1.61E-07
		% predicted FEV1 post-bd	0.70	4.01E-03
		% predicted FEV1 pre-bd	0.60	2.41E-05
		% predicted FVC post-bd	0.77	7.65E-04
		% predicted FVC pre-bd	0.68	6.05E-07
		(FEV1 / FVC) post-bd	-0.63	1.23E-02
		(FEV1 / FVC) pre-bd	-0.56	8.77E-05

	GSE47460	% predicted DLCO	0.78	2.14E-30
	GPL14550	% predicted FEV1 post-bd	0.57	5.54E-09
		% predicted FEV1 pre-bd	0.66	2.62E-21
		% predicted FVC post-bd	0.65	5.34E-12
		% predicted FVC pre-bd	0.75	5.45E-30
		(FEV1 / FVC) post-bd	-0.47	3.59E-06
		(FEV1 / FVC) pre-bd	-0.54	2.84E-13
	GSE32537	% predicted DLCO	0.20	4.69E-02
		% predicted FVC pre-bd	0.25	8.57E-03
AQP5	GSE47460	% predicted DLCO	-0.39	1.68E-02
	GPL6480	% predicted FEV1 post-bd	-0.60	1.92E-02
		% predicted FEV1 pre-bd	-0.32	3.81E-02
		% predicted FVC post-bd	-0.59	2.15E-02
		% predicted FVC pre-bd	-0.36	1.74E-02
	GSE47460	% predicted DLCO	-0.64	5.32E-18
	GPL14550	% predicted FEV1 post-bd	-0.38	3.10E-04
		% predicted FEV1 pre-bd	-0.46	1.04E-09
		% predicted FVC post-bd	-0.48	1.90E-06
		% predicted FVC pre-bd	-0.56	2.22E-14
COL15A1	GSE47460	% predicted DLCO	-0.59	1.16E-04
	GPL6480	% predicted FEV1 post-bd	-0.76	1.06E-03
		% predicted FEV1 pre-bd	-0.47	1.41E-03
		% predicted FVC post-bd	-0.79	4.49E-04
		% predicted FVC pre-bd	-0.56	7.93E-05
		(FEV1 / FVC) post-bd	0.64	9.92E-03
		(FEV1 / FVC) pre-bd	0.55	1.24E-04
	GSE47460	% predicted DLCO	-0.61	2.40E-16
	GPL14550	% predicted FEV1 post-bd	-0.45	8.92E-06
		% predicted FEV1 pre-bd	-0.52	1.60E-12
		% predicted FVC post-bd	-0.60	6.29E-10
		% predicted FVC pre-bd	-0.63	6.50E-19
		(FEV1 / FVC) post-bd	0.53	9.95E-08
		(FEV1 / FVC) pre-bd	0.51	9.33E-12
COL1A1	GSE47460	% predicted DLCO	-0.61	5.41E-05
	GPL6480	% predicted FEV1 post-bd	-0.89	7.94E-06
		% predicted FEV1 pre-bd	-0.55	1.54E-04
		% predicted FVC post-bd	-0.89	7.58E-06
		% predicted FVC pre-bd	-0.64	4.52E-06
		(FEV1 / FVC) pre-bd	0.58	4.88E-05
	GSE47460	% predicted DLCO	-0.65	1.78E-18
	GPL14550	% predicted FEV1 post-bd	-0.48	1.70E-06
		% predicted FEV1 pre-bd	-0.51	5.43E-12
		% predicted FVC post-bd	-0.63	3.31E-11
		% predicted FVC pre-bd	-0.62	3.42E-18
		(FEV1 / FVC) pre-bd	0.52	1.80E-12
COMP		% predicted DLCO	-0.73	3.83E-07

	GSE47460	% predicted FEV1 post-bd	-0.74	1.61E-03
	GPL6480	% predicted FEV1 pre-bd	-0.65	2.36E-06
		% predicted FVC post-bd	-0.80	3.85E-04
		% predicted FVC pre-bd	-0.68	4.42E-07
		(FEV1 / FVC) post-bd	0.62	1.44E-02
		(FEV1 / FVC) pre-bd	0.51	4.52E-04
	GSE47460	% predicted DLCO	-0.74	6.47E-26
	GPL14550	% predicted FEV1 post-bd	-0.53	9.12E-08
		% predicted FEV1 pre-bd	-0.58	1.23E-15
		% predicted FVC post-bd	-0.67	1.43E-12
		% predicted FVC pre-bd	-0.69	2.22E-23
		(FEV1 / FVC) post-bd	0.54	6.70E-08
		(FEV1 / FVC) pre-bd	0.56	2.85E-14
	GSE32537	% predicted DLCO	-0.25	1.56E-02
		% predicted FVC pre-bd	-0.27	3.77E-03
DCXR	GSE47460	(FEV1 / FVC) post-bd	-0.57	2.77E-02
	GPL6480	(FEV1 / FVC) pre-bd	-0.40	7.80E-03
	GSE47460	% predicted DLCO	0.31	1.18E-04
	GPL14550	% predicted FVC pre-bd	0.31	7.41E-05
		(FEV1 / FVC) post-bd	-0.28	7.95E-03
		(FEV1 / FVC) pre-bd	-0.32	3.27E-05
	GSE32537	% predicted DLCO	-0.27	9.32E-03
		% predicted FVC pre-bd	-0.23	1.55E-02
IL13RA2	GSE47460	% predicted DLCO	-0.53	7.30E-04
	GPL6480	% predicted FEV1 post-bd	-0.78	6.46E-04
		% predicted FEV1 pre-bd	-0.43	3.65E-03
		% predicted FVC post-bd	-0.83	1.13E-04
		% predicted FVC pre-bd	-0.53	2.91E-04
		(FEV1 / FVC) post-bd	0.70	3.50E-03
		(FEV1 / FVC) pre-bd	0.57	6.13E-05
	GSE47460	% predicted DLCO	-0.67	1.47E-20
	GPL14550	% predicted FEV1 post-bd	-0.51	4.92E-07
		% predicted FEV1 pre-bd	-0.52	1.30E-12
		% predicted FVC post-bd	-0.64	1.21E-11
		% predicted FVC pre-bd	-0.65	2.73E-20
		(FEV1 / FVC) post-bd	0.55	3.37E-08
		(FEV1 / FVC) pre-bd	0.55	9.91E-14
	GSE32537	% predicted FVC pre-bd	-0.26	5.69E-03
MMP7	GSE47460	% predicted DLCO	-0.65	1.27E-05
	GPL6480	% predicted FEV1 post-bd	-0.76	8.92E-04
		% predicted FEV1 pre-bd	-0.49	7.46E-04
		% predicted FVC post-bd	-0.76	1.00E-03
		% predicted FVC pre-bd	-0.54	1.72E-04
		(FEV1 / FVC) pre-bd	0.44	3.37E-03
	GSE47460	% predicted DLCO	-0.70	1.68E-22
	GPL14550	% predicted FEV1 post-bd	-0.46	6.39E-06

		% predicted FEV1 pre-bd	-0.58	1.35E-15
		% predicted FVC post-bd	-0.56	1.13E-08
		% predicted FVC pre-bd	-0.68	1.15E-22
		(FEV1 / FVC) pre-bd	0.53	1.05E-12
PAPSS2	GSE47460	% predicted DLCO	0.60	1.02E-04
	GPL6480	% predicted FEV1 post-bd	0.60	1.75E-02
		% predicted FEV1 pre-bd	0.50	5.87E-04
		% predicted FVC post-bd	0.64	1.10E-02
		% predicted FVC pre-bd	0.61	1.61E-05
		(FEV1 / FVC) pre-bd	-0.54	2.06E-04
	GSE47460	% predicted DLCO	0.70	6.82E-23
	GPL14550	% predicted FEV1 post-bd	0.50	7.95E-07
		% predicted FEV1 pre-bd	0.55	1.10E-13
		% predicted FVC post-bd	0.62	1.56E-10
		% predicted FVC pre-bd	0.67	8.12E-22
		(FEV1 / FVC) pre-bd	-0.56	3.04E-14

DLCO: diffusing capacity of the lungs for carbon monoxide; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; pre/post bd: pre/post bronchodilator

Table S7. *ENPP2* differentially methylated features between colorectal cancer (CRC) cells prior to and post to 5-Aza-CdR treatment (GSE51815). Negative $\Delta\beta$ values verify demethylation of all examined CGs. Adapted from (Panagopoulou, Fanidis et al. 2021).

CG ID	M β value 1*	M β value 2*	$\Delta\beta$ value#	Location relative to gene	FDR
HT116 1KO vs AZA treated					
cg00320790	0.785	0.595	-0.190	Body	1.9E-02
cg07236691	0.799	0.592	-0.207	Body	2.5E-03
cg09444531	0.391	0.288	-0.103	Body	6.1E-03
cg20048037	0.510	0.365	-0.146	Body	3.8E-02
cg20162626	0.543	0.412	-0.131	Body	2.5E-03
cg23725583	0.578	0.424	-0.154	Body	7.7E-05
cg26078665	0.825	0.749	-0.076	Body	8.5E-03
cg04452959	0.171	0.150	-0.021	TSS200	8.0E-03
cg06998282	0.708	0.458	-0.251	TSS1500	2.5E-03
cg14409958	0.463	0.332	-0.131	TSS1500	1.9E-02
cg02534163	0.468	0.181	-0.287	1st Exon	7.7E-05
HT116 3BKO vs AZA treated					
cg00320790	0.946	0.678	-0.268	Body	1.2E-03
cg01243251	0.930	0.812	-0.118	Body	3.2E-02
cg07236691	0.835	0.515	-0.319	Body	2.8E-04
cg09444531	0.617	0.403	-0.214	Body	2.0E-02
cg20048037	0.842	0.518	-0.325	Body	2.9E-04
cg20162626	0.752	0.463	-0.288	Body	2.0E-02
cg23725583	0.928	0.657	-0.270	Body	1.5E-04
cg26078665	0.873	0.752	-0.122	Body	7.3E-04
cg02709432	0.417	0.263	-0.154	TSS200	4.4E-04
cg04452959	0.581	0.347	-0.234	TSS200	1.1E-04
cg02156680	0.188	0.106	-0.082	TSS1500	6.3E-04
cg06998282	0.953	0.571	-0.381	TSS1500	2.0E-04
cg14409958	0.768	0.395	-0.373	TSS1500	2.2E-04
cg02534163	0.958	0.584	-0.374	1st Exon	2.2E-04
HT116 WT vs AZA treated					
cg00320790	0.873	0.668	-0.206	Body	5.6E-03

cg01243251	0.894	0.769	-0.125	Body	1.0E-04
cg07236691	0.815	0.610	-0.205	Body	9.6E-03
cg09444531	0.860	0.598	-0.262	Body	4.7E-03
cg20048037	0.654	0.543	-0.110	Body	9.6E-03
cg20162626	0.613	0.483	-0.130	Body	1.4E-02
cg23725583	0.774	0.615	-0.160	Body	3.2E-02
cg26078665	0.856	0.740	-0.116	Body	1.4E-02
cg02709432	0.447	0.361	-0.087	TSS200	3.4E-02
cg04452959	0.526	0.389	-0.138	TSS200	5.6E-03
cg06998282	0.887	0.618	-0.269	TSS1500	1.0E-03
cg14409958	0.705	0.474	-0.231	TSS1500	1.3E-02

*Mean β ($M\beta$) value 1 represents methylation without 5-Aza-CdR treatment and *Mean β ($M\beta$) value 2; methylation in 5-Aza-CdR treated cell lines; # $\Delta\beta$ value: (Mean β value 2-Mean β value 1); PA: Promoter Associated, TSS: Transcription Start Site, 5-Aza-CdR: 5-AZA-2'-deoxycytidine, HT116 1KO: colon cancer cell line, knockout of the DNA methyltransferase gene DNMT1, HT116 3BKO: colon cancer cell line, knockout of the DNA methyltransferase gene DNMT3B

Table S8. Taxa of the indicated taxonomic level and diet. Adopted from (Galaris, Fanidis et al. 2022).

Inter-tissue common taxa unique per dietary regime		
Taxonomic level	Unique in	Unique taxa
Phylum	HFD	Fusobacteria; Acidobacteria
	Ctrl	-
Family	HFD	Peptostreptococcaceae; Muribaculaceae; Rhodobacteraceae; Bacillaceae; Bacteroidaceae; Bifidobacteriaceae; Azospirillaceae; Family_XIII; Listeriaceae; Leuconostocaceae; Flavobacteriaceae; Beijerinckiaceae; Nocardiaceae; Dermacoccaceae; Mitochondria; Sphingobacteriaceae; Nocardiodaceae
	Ctrl	-
Inter-tissue common taxa		
Taxonomic level	Found in	Taxa
Phylum	HFD	Bacteroidetes; Proteobacteria; Firmicutes; Actinobacteria; Cyanobacteria
	Ctrl	Firmicutes; Actinobacteria; Proteobacteria; Bacteroidetes; Cyanobacteria; Fusobacteria; Acidobacteria
Family	HFD	Propionibacteriaceae; Staphylococcaceae; Streptococcaceae; Xanthobacteraceae; Peptoniphilaceae; Corynebacteriaceae; Lactobacillaceae; Micrococcaceae; Pasteurellaceae; Caulobacteraceae; Moraxellaceae; Peptostreptococcaceae; Rhizobiaceae; Prevotellaceae; Lachnospiraceae; Carnobacteriaceae; Neisseriaceae; Actinomycetaceae; Muribaculaceae; Veillonellaceae; Porphyromonadaceae; Burkholderiaceae; Weeksellaceae; Aeromonadaceae; Rhodobacteraceae; Erysipelotrichaceae; Enterobacteriaceae; Ruminococcaceae; Atopobiaceae; Aerococcaceae; Bacillaceae; Fusobacteriaceae; Sphingomonadaceae; Microbacteriaceae; Bacteroidaceae; Bifidobacteriaceae; Azospirillaceae; Pseudomonadaceae; Family_XIII; Listeriaceae; Xanthomonadaceae; Leuconostocaceae; Flavobacteriaceae; Beijerinckiaceae; Nocardiaceae; Dermacoccaceae; Mitochondria; Sphingobacteriaceae; Nocardiodaceae
	Ctrl	Xanthobacteraceae; Rhizobiaceae; Caulobacteraceae; Staphylococcaceae; Propionibacteriaceae; Streptococcaceae; Peptoniphilaceae; Erysipelotrichaceae; Corynebacteriaceae; Lactobacillaceae; Burkholderiaceae; Pasteurellaceae; Micrococcaceae; Carnobacteriaceae; Lachnospiraceae; Prevotellaceae; Neisseriaceae; Enterobacteriaceae; Moraxellaceae; Porphyromonadaceae; Veillonellaceae; Atopobiaceae; Actinomycetaceae; Weeksellaceae; Ruminococcaceae; Sphingomonadaceae; Microbacteriaceae; Xanthomonadaceae; Aerococcaceae; Pseudomonadaceae; Aeromonadaceae; Fusobacteriaceae

Table S9. Number of taxa shared between conditions as revealed by pairwise comparisons. Adopted from (Galaris, Fanidis et al. 2022).

		GutCtrl	GutHFD	LiverCtrl	LiverHFD	LungCtrl	LungHFD
Phylum	GutCtrl	-	5	5	5	5	5
	GutHFD	5	-	7	7	6	8
	LiverCtrl	5	7	-	7	6	7
	LiverHFD	5	7	7	-	6	8
	LungCtrl	5	6	6	6	-	6
	LungHFD	5	8	7	8	6	-
Family	GutCtrl	-	41	34	39	34	36
	GutHFD	41	-	45	52	38	50
	LiverCtrl	34	45	-	53	40	55
	LiverHFD	39	52	53	-	41	59
	LungCtrl	34	38	40	41	-	43
	LungHFD	36	50	55	59	43	-

Table S10. Detected genera of the recorded phyla and families. Red font marks genera affected by high fat diet in at least a single tissue (HFD to control relative abundance difference other than zero). Adopted from (Galaris, Fanidis et al. 2022).

Phylum	Family	Genus
Firmicutes	Streptococcaceae	<i>Streptococcus</i>
		<i>Lactococcus</i>
	Staphylococcaceae	<i>Staphylococcus</i>
		<i>Salinicoccus</i>
		<i>Jeotgalicoccus</i>
		<i>Gemella</i>
		<i>Peptoniphilus</i>
	Peptoniphilaceae	<i>Anaerococcus</i>
		<i>Finegoldia</i>
		<i>Gallicola</i>
		<i>Parvimonas</i>
		<i>Tissierella</i>
		<i>Ezakiella</i>
		<i>Murdochiella</i>
		<i>W5053</i>
<i>Sedimentibacter</i>		
Proteobacteria	Pasteurellaceae	NA

Table S11. Super-antigen related species detected in our dataset. Reference list of related species retrieved from UniProt database and/or the literature. Adopted from (Galaris, Fanidis et al. 2022).

Phylum	Family	Species	Source
Firmicutes	Peptoniphilaceae	<i>Finegoldia magna</i>	Literature
	Streptococcaceae	<i>Lactobacillus delbruecki</i>	TrEBML
		<i>Lactococcus lactis</i>	TrEBML
	Staphylococcaceae	<i>Staphylococcus aureus</i>	KB; TrEBML; Literature
		<i>Staphylococcus massiliensis</i>	TrEBML
		<i>Staphylococcus epidermis</i>	TrEBML
Proteobacteria	Moraxellaceae	<i>Acinetobacter baumannii</i>	TrEBML

Table S12. *IPF_vs_Ctrl* and *Bleomycin_vs_Ctrl* Fibromine datasets interrogated for *LCN2* differential transcription. It was found up-regulated in several occasions. Red font marks comparisons iconized in figures 21 and 23 of the main text. Adopted from (Galaris, Fanidis et al. 2023).

Species	PMID	Summary	Dataset	Platform	# (Exper/ Control)	<i>LCN2</i> FC	FDR
<i>Homo sapiens</i>	26560100	IPF lung samples were compared to healthy ones collected during an exploratory surgery.	GSE47460	GPL6480	28/15	4.50	1.17E-05
	23783374	LTRC pulmonary fibrosis samples transcriptome was compared to that of healthy controls.	GSE47460	GPL14550	84/75	3.86	9.48E-19
	19363140	Lungs of patients collected during pulmonary transplant compared with control ones derived from cancer free lung margins.	GSE32537	GPL6244	115/44	2.28	2.46E-08
	30111332	Fresh frozen lung samples from IPF patients were compared to normal lung tissue flanking pulmonary cancer areas.	GSE10667	GPL4133	21/14	3.58	9.10E-03
	21360508	Microarray data from IPF patients were juxtaposed to data from control individuals.	GSE110147	GPL6244	22/11	2.33	1.53E-03
	25217476	Differences between IPF and control transcriptomic profile were examined based on lung biopsies or explants.	GSE48149	GPL16221	12/5	6.15	2.32E-02
	28230051	LTRC IPF and control samples were examined to identify transcriptional differences between disease and steady state conditions.	GSE53845	GPL6480	39/7	5.43	3.97E-04
	16998095	LTRC IPF and control samples were examined to identify transcriptional differences between disease and steady state conditions.	GSE92592	GPL11154	19/18	6.77	5.69E-11
	28081703	Sporadic idiopathic interstitial pneumonia samples were compared to control samples.	GSE5774	GPL4255	13/17	-1.85	1.3E-01
	29329637	RNA-seq assessed the transcriptome of IPF and control FFPE lung samples.	GSE83717	GPL11154	6/5	2.83	1.4E-01
	21974901	IPF samples were compared with healthy controls using RNA-seq.	GSE99621	GPL16791	8/8	1.75	1.7E-01
	24647608	Healthy samples obtained during lung transplantation were compared to IPF.	GSE24206	GPL570	8/6	3.36	2.0E-01
	21241464	mRNA of IPF and control individuals was examined using next generation sequencing.	GSE52463	GPL11154	6/4	2.25	2.3E-01
	26453058	LRTC IPF and uninvolved lung samples of cancer and a transplant patient were compared. Control samples from patients with primary spontaneous pneumothorax was compared to samples originating from IPF patients.	GSE21369	GPL570	11/6	1.92	4.3E-01
	23648810	Total RNA was collected from C57BL/6 mouse lungs treated or not with bleomycin. (BlmD14_vs_Ctrl)	GSE72073	GPL17586	4/3	1.44	4.6E-01
	<i>Mus musculus</i>	21602491	Mice treated for 21 days with bleomycin were compared to saline treated control ones. (BlmD21_vs_Ctrl)	GSE37633	GPL6885	4/6	5.35
19966781		Mice receiving bleomycin via oropharyngeal instillation were compared to control counterparts. (BlmD14_vs_Ctrl)	GSE25640	GPL1261	3/3	3.03	3.25E-02
23459460		Affymetrix microarrays were used to assess murine transcriptome 14 days post to bleomycin instillation compared to control untreated samples. (BlmD14_vs_Ctrl)	GSE18800	GPL1261	3/4	4.22	2.67E-04
23565148		Bleomycin-treated mice were compared to PBS-treated ones to assess differences in their transcriptomic profiles. (BlmD14_vs_Ctrl)	GSE34814	GPL13912	3/3	3.01	2.32E-03
19652365		Male C57BL/6J mice received bleomycin or saline treatment. (BlmD14_vs_Ctrl)	GSE40151	GPL1261	16/16	-1.66	1.89E-02
-		Mice receiving bleomycin via oropharyngeal instillation were compared to control counterparts. (BlmD14_vs_Ctrl)	GSE16846	GPL339	3/3	4.29	3.42E-02
-		Mice received bleomycin or water via tracheal injection. (BlmD21_vs_Ctrl)	GSE77326	GPL15887	6/6	1.41	3.47E-02

LTRC: lung tissue research consortium; Dotted line separates significant (top) from non-significant (bottom) human comparisons.

Table S13. Single cell datasets used for the exploration of *LCN2/Lcn2* cell specific expression. Adapted from (Galaris, Fanidis et al. 2023).

PMID	Summary	Dsataset	Comparison	# Exp / Control	# Cells	# Genes
33650774	Transcriptome examination of IPF or control origin single cells based on the integration of three independent cohorts.	github	IPF_vs_Ctrl	19/29	181814	49498
32832599	Transcriptome interrogation of whole lung dissociates at the single cell level.	GSE136831	IPF_vs_Ctrl	32/28	243472	45947
32832598	Single cell exploration of non-fibrotic and pulmonary fibrosis lung samples.	github	IPF_vs_Ctrl	12/10	89326	33694
32678092	Time-course experiment of bleomycin-treated and control mice at a single cell resolution.	github	Blm_vs_Ctrl	21/7	29142	23400

Table S14. Neutrophil and epithelial cells of multiple human and mouse organs are marked by *ENPP2/Enpp2* expression (CellMarker2.0). Adopted from (Galaris, Fanidis et al. 2023).

Species	PMID	Tissue	Cell type	Evidence
<i>Homo sapiens</i>	34128959	Blood	Neutrophil	Experiment
	31753849	Colon	Progenitor cell	Experiment
	31604275	Kidney	Pelvic epithelial cell	Experiment
	33730555	Lacrimal gland	Ductal cell	Experiment
	33619773	Liver	Cholangiocyte	Experiment
	34721400	Peripheral blood	Polymorphonuclear myeloid-derived suppressor(PMN-MDSC) cell	Experiment
	35111153	Peripheral blood	Neutrophil progenitor cell	Experiment
<i>Mus musculus</i>	32424270	Blood	Granulocytic neutrophil	Experiment
	33028096	Blood vessel	Fibroblast	Experiment
	35175542	Blood vessel	Neutrophil	Experiment
	29466336	Bone marrow	Neutrophil	Single-cell sequencing
	29915358	Bone marrow	Stage I neutrophil	Single-cell sequencing
	32221280	Brain	Choroid cell	Experiment
	32640942	Brain	Neutrophil	Experiment
	33028409	Brain	Microglial cell	Experiment
	32759216	Fetal ovary	Blood-related cell	Experiment
	29622724	Kidney	Neutrophil	Single-cell sequencing
	34367152	Kidney	Neutrophil	Experiment
	24739965	Lung	Type II pneumocyte	Single-cell sequencing
	34389631	Lung	Neutrophil	Experiment
	34301765	Meninge	Neutrophil	Experiment
	32059779	Testis	Endothelial cell	Experiment

Table S15. *TKS5* differential transcription between *IPF* and control samples from *Fibromine*. Adopted from (Barbayaanni, Kanelloupolou et al. 2023).

PMID	Samples	GEO dataset	Platform	IPF/Ctrl	FC	FDR
19363140 https://www.ncbi.nlm.nih.gov/pubmed/19363140	Lungs of patients collected during pulmonary transplant compared with control ones derived from cancer free lung margins.	GSE10667 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE10667	GPL4133 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL4133	21/14	1.29	3.43E-02
30111332 https://www.ncbi.nlm.nih.gov/pubmed/30111332	Fresh frozen lung samples from IPF patients were compared to normal lung tissue flanking pulmonary cancer areas.	GSE110147 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE110147	GPL6244 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL6244	22/11	-1.23	1.95E-03
21241464 https://www.ncbi.nlm.nih.gov/pubmed/21241464	IPF lung tissue originated from LTRC were assessed in comparison to healthy lung tissue of cancer patients and a lung transplant.	GSE21369 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21369	GPL570 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL570	11/6	2.09	4.81E-02
21974901 https://www.ncbi.nlm.nih.gov/pubmed/21974901	IPF lung samples transcriptional profile was compared to that of control samples from lung volume reduction during transplantation.	GSE24206 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE24206	GPL570 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL570	8/6	2.10	1.08E-02
26560100 https://www.ncbi.nlm.nih.gov/pubmed/26560100	IPF lung samples were compared to healthy ones collected during an exploratory surgery.	GSE47460 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE47460	GPL6480 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL6480	28/15	1.23	3.69E-03
26560100 https://www.ncbi.nlm.nih.gov/pubmed/26560100		GSE47460 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE47460	GPL14550 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL14550	84/75	1.31	1.54E-20
21360508 https://www.ncbi.nlm.nih.gov/pubmed/21360508	Microarray data from IPF patients were juxtaposed to data from control individuals.	GSE48149 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48149	GPL16221 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL16221	12/5	1.56	2.79E-03
25217476 https://www.ncbi.nlm.nih.gov/pubmed/25217476	Differences between IPF and control transcriptomic profile were examined based on lung biopsies or explants.	GSE53845 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53845	GPL6480 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL6480	39/7	1.28	1.24E-02
28230051 https://www.ncbi.nlm.nih.gov/pubmed/28230051	LTRC IPF and control samples were examined to identify transcriptional differences between disease and steady state conditions.	GSE92592 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE92592	GPL11154 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL11154	19/18	1.78	4.24E-06
29329637 https://www.ncbi.nlm.nih.gov/pubmed/29329637	Scarred regions from explanted lungs were compared to healthy control lung tissues.	GSE99621 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99621	GPL16791 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL16791	8/8	1.52	3.70E-03

LTRC: lung tissue research consortium

Table S16. Demographic and clinical description of lung tissue donors. Adopted from (Barbayianni, Kanellopoulou et al. 2023).

Characteristic	IPF (n=20)	Control (n=9)	COPD (n=19)
Age (yr)(Mean ± SD)	64.8 ± 8.5	68.8 ± 14.8	67.9 ± 14.5
Sex, n (%)			
Males	12 (60%)	5 (55.6%)	15 (79%)
Females	8 (40%)	4 (44.4%)	4 (21%)
Race, n (%)			
White	19 (95%)	9 (100%)	19 (100%)
Hispanic	1 (5%)	0	0
Pulmonary function tests (Mean ± SD)			
FVC%	65.8 ± 14.1	93.9 ± 16.8	76.5 ± 14.7
DLCO%	45.6 ± 14.9	85.6 ± 13.8	58.9 ± 21.7
FEV1%	73.3 ± 14	92.6 ± 12.2	53.3 ± 21.7

FVC%: %predicted forced vital capacity; DLCO%: %predicted carbon monoxide diffusing capacity; FEV1%: %predicted forced expiratory volume in 1 second.

Table S17. Demographic and clinical description of lung fibroblast donors. Comparison between IPF and control groups was performed with two-sided unpaired t-test (**p=0.0029 ****p<0.0001). Adopted from (Barbayianni, Kanellopoulou et al. 2023).

Characteristic	IPF (n=5)	Control (n=5)
Age (yr)(Mean ± SD)	66.7± 2.9	64.8 ± 10.3
Sex, n (%)		
Males	3 (60%)	3 (60%)
Females	2 (40%)	2 (40%)
Pulmonary function tests (Mean ± SD)		
DLCO%	47.5 ± 12.4	52 ± 12.4
FEV1/FVC%	96 ± 7**	72 ± 10
Podosomes		
% Cells with podosomes/area	51.6 ± 5.7****	13.4 ± 2.5
Podosomes/cell	2.9±0.4****	0.8±0.3

FVC%: %predicted forced vital capacity; DLCO%: %predicted carbon monoxide diffusing capacity; FEV1%: %predicted forced expiratory volume in 1 second.

Table S18. Factors that create a transcriptional profile similarly to that of TGFβ-induced *Tks5*^{+/-} lung fibroblasts. Adopted from (Barbayianni, Kanellopoulou et al. 2023).

NCS	Name	Main Target	Targets	Cell line	PMID	IPF Relevance
1.760	GSK-1070916	Aurora kinase	AURKB AURKC AURKA CYP2D6 CYP3A4	BT20	32761869 https://pubmed.ncbi.nlm.nih.gov/32761869/	Inhibition of Aurora Kinase B attenuates fibroblast activation and pulmonary fibrosis
1.733	PSB-069	NTPDase	ENTPD1 ENTPD2 ENTPD3	MCF7	-	-
1.716	SANT-2	Smoothed receptor	SMO DHH IHH PTCH1 SHH	HA1E	-	-
1.697	entinostat	HDAC	HDAC1 HDAC2 HDAC3 HDAC9	YAPC	28315487 https://pubmed.ncbi.nlm.nih.gov/28315487/	Reversion of TGFβ1-induced, XPLN/ SPARK-mediated extracellular matrix turnover in human foetal lung fibroblast cell line
1.695	pazopanib	VEGFR KIT PDGFR	KDR KIT FLT1 FLT4 PDGFRB PDGFR BRAF CYP2B6 CYP2C8 CYP2E1 DDR2 CSF1R FGF1 FGFR1 FGFR3 ITK SH2B3	MCF10A	21992121 https://pubmed.ncbi.nlm.nih.gov/21992121/ 27279371 https://pubmed.ncbi.nlm.nih.gov/27279371/	Tyrosine kinase inhibitor attenuated worsening of lung function, maintained quality of life and reduced acute exacerbations of IPF patients PDGFR inhibition attenuated enhanced differentiation and proliferation of myofibroblasts
1.683	crizotinib	ALK	ALK MET CYP2B6 CYP3A5 MST1R ROS1	TMD8	15563636 https://pubmed.ncbi.nlm.nih.gov/15563636/	PDGFβ1-induced lung fibrosis is blocked by an orally active ALK5 kinase inhibitor
1.671	emetine	Protein synthesis	RPS2	A375	18503048 https://pubmed.ncbi.nlm.nih.gov/18503048/	Increase of Sulf1 mRNA levels, a potential negative regulator of TGFβ-1 induced fibrogenesis, through inhibition of protein translation
1.667	TAK-875	Insulin	FFAR1 INS	A375	-	-
1.664	filgotinib	JAK	JAK1 JAK2 JAK3 TYK2	XC.L10	29440315 https://pubmed.ncbi.nlm.nih.gov/29440315/	Block of endothelial to mesenchymal transition and artery smooth muscle cell to myofibroblast transition via JAK inhibition thus reducing bleomycin-induced pulmonary fibrosis
1.664	arofylline	Phosphodiesterase	PDE4A PDE4B PDE4C PDE4D	PC3	-	-
1.663	raloxifene	Estrogen receptor	ESR1 ESR2 ACVRL1 ENG	HT29	35841004 https://pubmed.ncbi.nlm.nih.gov/35841004/	Reduction of systemic sclerosis (PSC)-derived fibroblast proliferation, production of extracellular matrix and skin fibrosis in human organoids and mouse model.
1.661	WH-4023	Src	LCK SRC ABL1	MCF10A	38998281 https://pubmed.ncbi.nlm.nih.gov/38998281/	Saracatinib, a Selective Src Kinase Inhibitor, Blocks Fibrotic Responses in Preclinical Models of Pulmonary Fibrosis
1.661	brefeldin-a	Protein synthesis BIG1	ARFGEF1 ARFGEF2 GBF1 ARF1 CYTH2	HCT116	8429043 https://pubmed.ncbi.nlm.nih.gov/8429043/	Inhibition of intracellular degradation and thus secretion of collagen in normal human fetal lung fibroblasts
1.655	TGX-221	PI3K	PIK3CB PIK3CD	MCF10A	2184893 https://pubmed.ncbi.nlm.nih.gov/2184893/	PI3K inhibition attenuates myofibroblast differentiation in the human lungs
1.654	motesanib	KIT PDGFR VEGFR	FLT1 FLT4 KDR KIT PDGFR RET	MCF10A	-	-

NCS: normalized connectivity score

Table S19. Description of single cell datasets used. Adapted from (Barbayianni, Kanellopoulou et al. 2023).

Disease	PMID	Dataset	Tissue	Cells	#Disease/ #Ctrl	Note
COVID-19	32591762	figshare	NS	135600	27/5	Main object
	33361824	figshare		88177	32/16	
	32514174	covid cell atlas	PBMC	44721	7/6	
	32810438	fastgenomics		99049	27/22	Cohort one
	32398875	UCSC cell browser	BALF	66452	9/4	
	33429418	GSE155249		77146	15/0	Main object
	33257409	GSE158127	Lung	155413	12/10	Bharat samples
33915569	SCP1052	106792		16/0		
ILDs*	33650774	github		233780	33/29	Integrated object

Idiopathic pulmonary fibrosis; Hypersensitivity pneumonitis; Interstitial lung disease associated with systemic sclerosis; Myositis; Nonspecific interstitial pneumonia; Unclassifiable; Chronic hypersensitivity pneumonitis; Sarcoidosis; Extrinsic allergic alveolitis. #NS: nasopharyngeal samples

Figure S20. Stimuli used for the treatment of HCK-8 cells. Adapted from (Magkrioti, Antonopoulou et al. 2022).

Stimulus	Target C (ng/ml)	Stimulus	Target C (ng/ml)
Pam3CSK4	1000	Activin A	100
Poly(I:C) HMW	10000	DKK-1	100
LPS-EK standard	10000	IFN beta	50
FLA-ST standard	1000	IL-12	50
FSL1	1000	IL17F	100
Imiquimod	1000	IL22	200
ssRNA40/LyoVec	1000	IL23	100
C12-iE-DAP	1000	PDGF-AA	50
MDP	10000	PDGF-AB	100
Tri-DAP	10000	PDGF-BB	50
IL17A	100	IL16 (121)	100
NT-3	200	IL6	100
NT-4	100	TNFA	100
SCF	100	IL-13	100
SCGF-A	500	CXCL14/BRAK	200
SFAS ligand	10	CCL4L1/LAG1	200
Soluble RANK ligand	100	CCL3L1/LD78B	200
TGFB1	10	CCL28/MEC	50
TGFB2	10	CCL19/MIP3B	500
TGFB3	10	Amphiregulin (98)	100
VCAM-1	100	BMP-13/CDMP-2	300
VEGF 121	100	BMP-3	100
WISP-3	150	CTGF	50
ANG-1	250	CTGFL/WISP-2	50
ANG-2	500	CYR61	200
GMF-beta	100	EG-VEGF	100
CCL2	200	Endostatin	1000
CCL3	200	Epigen	400
CCL4	1000	FGF-10	500
CCL5	500	FGF-16	100
SDF-1A	200	FGF-17	100
SDF-1B	200	FGF-18	100
GM-CSF	20	FGF-21	10
IFN-G	100	FGF-23	50
IL-1A	50	FGF-4	100
IL-1B	50	FGF-5	50
IL-10	10	FGF-6	250
Human EGF (Animal Free)	100	FGF-8	100
Beta-NGF	50	FGF-9	100

FGF-acidic	100	Pleiotrophin	100
Galectin-1	100	Slit2-N	500
Galectin-3	20	APRIL	200
GDF-11	10	BMP-2	100
IL4	100	BMP-4	200
HGF	50	BMP-6	300
BDNF	50	BMP-7	100
Klotho	50	FGF-20	20
NOGGIN	200	G-CSF	100
OPG	10	GDF-7	100
PDGF-CC	125	IL-11	100
PEDF	200	WNT-1	50
sCD100	50	Betacellulin	50
sCD14	130	IL-15	100
sCD40 ligand	500	IL19	100
SDLL-1	500	IL20	200
SDLL-4	500	IL9	30
SFRP-1	100	KGF	100
Sonic HedgeJoa	200	IL17B	100
SPARC	100	IL17D	200
TGFA	200	IL17E	100
CNTF	200	Dexamethasone	392
GDNF	200	Riluzole hydrochloride	4006
TRAIL	100	Norethindrone	3999
TWEAK	100	Clomipramine hydrochloride	4005
VEGF 165	100	Clenbuterol hydrochloride	4015
VEGF-B	100	Betaxolol hydrochloride	3989
Visfatin	250	Bisacodyl	3975
WISP-1	200	Cisapride	4007
WNT-7A	25	Digitoxigenin	3970
HB-EGF	50	Progesterone	4025
IGF-BP2	500	Serotonin hydrochloride	5901
IGF-BP4	100	Benperidol	3967
IGF-BP5	100	Betahistine mesylate	5648
IGF-BP7	30	Hexestrol	4001
IGF-II	100	Tolmetin sodium salt dihydrate	3519
CDNF	10	Carbachol	3982
MANF	100	Promethazine hydrochloride	3979
Midkine (MK)	100	Nisoxetine hydrochloride	3528

Menadione	3995
Prestw-1494	
Hydrocotarnine	3989
Procyclidine hydrochloride	3564
Digoxin	4061
Metacycline	3716
Beta-Escin	3620
Epigallocatechin-3-gallate (EGCG)	45840
Mepyramine	2854
Lovastatin	3959
High glucose	9008000
Forskolin / Coleonol	4105
Staurosporine	466
PMA (Phorbol 12-myristate 13- acetate)	500
Insulin	1000
Formaldehyde	150
NaCl	8766000
H2O2	17005
Dimethyloxalyglycine	175140

Original publications

Fibromine is a multi-omics database and mining tool for target discovery in pulmonary fibrosis

Fibromine: μια πολυ-ωματική βάση δεδομένων και εργαλείο εξόρυξης δεδομένων για την ανακάλυψη στόχων στην πνευμονική ίνωση

Η ιδιοπαθής πνευμονική ίνωση είναι μια θανάσιμη ινοπολλαπλασιαστική ασθένεια με περιορισμένες θεραπευτικές επιλογές. Η εξέταση διαφορικής έκφρασης επηρεασμένων περιοχών υπήρξε καταλυτική για την ανατομία εμπλεκόμενων παθογενετικών μηχανισμών και την ανακάλυψη θεραπευτικών στόχων. Εντούτοις, οι προσπάθειες σύγκρισης/εξόρυξης δεδομένων από τα πολλαπλά σχετικά δημοσίως διαθέσιμα σύνολα δεδομένων με στόχο την επαλήθευση και τον σχηματισμό ερευνητικών υποθέσεων είναι πολύ λίγες. Σε αυτό το πλαίσιο παρουσιάζουμε το Fibromine, μία ενωποιητική βάση δεδομένων και περιβάλλον εξερεύνησης που αποτελείται από συνεκτικώς επαναλυμένα, χειρωνακτικώς σχολιασμένα μεταγραφωμικά και πρωτεωμικά σύνολα δεδομένων πνευμονικής ίνωσης. Τα σύνολα αυτά καλύπτουν ένα μεγάλο εύρος πειραματικών σχεδιασμών τόσο σε δείγματα ασθενών όσο και σε ζωικά μοντέλα. Πρόσβαση στο Fibromine δίνεται μέσω μιας R Shiny εφαρμογής (<http://www.fibromine.com/Fibromine>) η οποία προσφέρει λειτουργίες δυναμικής εξερεύνησης και ενωποίησης δεδομένων της βάσης σε πραγματικό χρόνο. Επιπροσθέτως, εισαγάγουμε ένα καινούργιο σύστημα συγκριτικής αξιολόγησης μεταγραφωμικών συνόλων δεδομένων χρήσει διαφόρων λανθανόντων χαρακτηριστικών τους. Μέσω αυτού καταλήγουμε σε βαθμονόμηση των συνόλων προκειμένου να βοηθήσουμε τον χρήστη στην επιλογή των καταλληλότερων εξ αυτών. Ακόμη, η κυτταρική εξειδίκευση στην έκφραση γονιδίων μπορεί να οπτικοποιηθεί και να εξερευνηθεί σε σύνολα δεδομένων έκφρασης μεμονωμένων κυττάρων σε μία προσπάθεια σύνδεσης δεδομένων κληρονομιάς (legacy data) με δεδομένα τεχνολογίας αιχμής. Τέλος, παρουσιάζονται διάφορα παραδείγματα χρήσης του εργαλείου, αναπαραγώγιμα από οποιονδήποτε μη εξειδικευμένο χρήστη, ο οποίος αποτελεί και τον πρωταρχικό αποδέκτη αυτού του εγχειρήματος.



OPEN

Fibromine is a multi-omics database and mining tool for target discovery in pulmonary fibrosis

Dionysios Fanidis¹, Panagiotis Moulos²✉ & Vassilis Aidinis¹✉

Idiopathic pulmonary fibrosis is a lethal lung fibroproliferative disease with limited therapeutic options. Differential expression profiling of affected sites has been instrumental for involved pathogenetic mechanisms dissection and therapeutic targets discovery. However, there have been limited efforts to comparatively analyse/mine the numerous related publicly available datasets, to fully exploit their potential on the validation/creation of novel research hypotheses. In this context and towards that goal, we present Fibromine, an integrated database and exploration environment comprising of consistently re-analysed, manually curated transcriptomic and proteomic pulmonary fibrosis datasets covering a wide range of experimental designs in both patients and animal models. Fibromine can be accessed via an R Shiny application (<http://www.fibromine.com/Fibromine>) which offers dynamic data exploration and real-time integration functionalities. Moreover, we introduce a novel benchmarking system based on transcriptomic datasets underlying characteristics, resulting to dataset accreditation aiming to aid the user on dataset selection. Cell specificity of gene expression can be visualised and/or explored in several scRNA-seq datasets, in an effort to link legacy data with this cutting-edge methodology and paving the way to their integration. Several use case examples are presented, that, importantly, can be reproduced on-the-fly by a non-specialist user, the primary target and potential user of this endeavour.

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive idiopathic pulmonary disease mainly manifested in older adults and characterized by extensive fibrosis of the lung interstitium that irreversibly affects normal lung function¹. Although presenting a highly heterogeneous clinical course, half of the IPF patients succumb to respiratory failure or life-threatening comorbidities within 2–5 years post diagnosis^{2–4}. Its enhanced morbidity, increased incidence among elder individuals⁴ and the lack of a curative option² render IPF a rare disease of major concern, especially for ageing societies. Currently, new studies focus on providing healthcare practitioners with invaluable information about patient stratification and prioritization, non-pharmaceutical treatment options and medications confronting potentially fatal comorbidities². Nevertheless, the task of new disease targets discovery is still greatly unfulfilled, as the most currently approved anti-fibrotics, nintedanib and pirfenidone, may delay lung function decline, but cannot help patients evade a fatal outcome⁵.

Since their first appearance, omics technologies have been extensively used to assess pathological deregulation in multiple molecular levels, thus physically leading to the progressive accumulation of a great volume of publicly available datasets. As far as IPF is concerned, a large number of expression profiling studies in IPF patients, as well as animal models, have been performed providing the scientific society with important lists of implicated targets such as that of Kim et al.⁶. Most of their datasets have been deposited in public repositories, but the qualitative assessment and mining of this vast legacy information, require significant expertise and time investment, not always available at the average research lab.

Irrespective of its laborious nature, comparative meta/re-analysis of publicly available datasets would provide additional means to discover novel pathogenic targets. Towards this direction, several attempts have been made to integrate IPF datasets^{7–11}, yet none of them spans a comprehensive collection of sample types, technologies and pathology models, while real-time data exploration features are either limited or absent.

Recognizing the need for a centralized, easily operable and comprehensive source of IPF(-related) data, we here introduce Fibromine, an open source application and database of integrable omics datasets accompanied by rich gene/RNA/protein level annotation and dataset meta-data. Currently, Fibromine hosts 60 consistently re-analysed and manually curated transcriptomic and proteomic datasets ranging over 42 unique comparisons,

¹Institute for Bioinnovation, Biomedical Sciences Research Center "Alexander Fleming", 16672 Athens, Greece. ²Institute for Fundamental Biomedical Research, Biomedical Sciences Research Center "Alexander Fleming", 16672 Athens, Greece. ✉email: moulos@fleming.gr; aidinis@fleming.gr

two species and several cell culture-based experiments. Moreover, and to increase data resolution, our database hosts a collection of more than 200,000 single cells originating from healthy and various pulmonary diseases samples. Fibromine can be accessed through a website (<http://www.fibromine.com/Fibromine>) that offers, amongst others: (i) within and across species integration of the supported datasets (*Dataset explorer*), (ii) interrogation of coding and non-coding gene expression patterns (*Gene explorer*; *miRNA explorer*), and (iii) exploration of protein abundance changes during disease (*Protein explorer*). In addition, other functionalities include the creation of disease-specific protein–protein interaction (PPI) (*Protein explorer*) and lung-specific gene co-expression networks (*Gene co-expression*), inter-connection with single cell resolution data visually via *Gene explorer* and in more detail via *Single cell data* tab, pathway analysis of consensus differentially expressed genes (cDEGs) and creation of heatmaps and volcano plots (*Dataset explorer*). Last but not least, in order to guide the user through transcriptomic datasets' underlying characteristics and inherent variability, a dataset benchmarking strategy based on seven distinct metrics has been applied (*Datasets benchmarking* tab). Conclusively, Fibromine not only aims to facilitate specialist and non-specialist users to validate their findings/observations and/or form nascent hypotheses prior to any time-consuming wet lab validation, but also to boost downstream research attempts via unconstrained access to all database data and integration/mining outputs.

Results

Fibromine datasets selection, re-analysis, curation and organization. Aiming to create a centralized repository for IPF omics data, we catalogued a plethora of publicly available microarray and RNA-seq transcriptomic datasets via PubMed and omicsdi.org using “IPF” and “bleomycin” as search keywords for human and mouse datasets, respectively. In the case of omicsdi.org, search results were further narrowed down using the “Transcriptomics” filter. Datasets included in the work of Villaseñor-Altamirano¹¹ were also taken into consideration and finally, references from an IPF transcriptomics review³ along with results from ReGEO¹² search using the “IPF” keyword were intersected with the already found datasets to form an initial pool of IPF datasets. To maintain the most informative, we excluded, among others, those consisting of less than three biological replicates, those of poor data quality (see further down), as well as those created using custom developed platforms (Fig. 1a). A few exceptions to the latter criterion were made to include datasets such as GSE31934 that address rarely explored scientific questions.

A major issue in all meta-analytic efforts that include data from various technological principles is the presence of technical heterogeneity within and across datasets that can, if ignored, greatly affect the data integration results by introducing non-biological sources of variability. Hence, to address technical variability in a scalable fashion that would facilitate future Fibromine updates without the requirement of already included data re-processing, we have re-analysed the collected transcriptomic datasets in a modular fashion maintaining the same re-analysis methods/parameters both across and within technologies in the highest degree possible. Then, to pinpoint and remove any sample outliers, all re-analysed datasets were manually curated using Principal Component Analysis (PCA)/MultiDimensional Scaling (MDS) and samples hierarchical clustering plots (Fig. 1a, Supplementary Fig. S1). Datasets failing to pass quality control criteria were excluded.

Subsequently, in order to further increase data resolution, we have pinpointed marker genes of all pathology-originating cells compared to healthy donated ones, as integrated by Mayr et al.¹³ Pathology versus healthy state comparisons were performed for over 40 cell types' top variable genes, resulting in an expression profiling of ten different pulmonary diseases.

Furthermore, in order to complement differential gene expression data, we retrieved deregulated proteins between IPF and control/healthy individuals from respective publications (Fig. 1a). The latter were selected via an extensive manual literature search and the only criterion applied for dataset selection was the provision of Differential Expression Analysis (DEA) results for more than just a couple of proteins in a tabular format. Proteomic datasets selection procedure was not as strict as the one for gene expression data, due to the relatively small number of the former publications relative to the latter^{14,15}. Proteins quantified in mixtures/same aptamer were removed, while DEA thresholds were kept *as is* in each original publication.

With the purpose of organising the aforementioned data, we created Fibromine, a dedicated application backed by a manually curated database that holds both DEA results and normalized transcriptomics data from 47 human and 13 murine datasets (Fig. 1a, Supplementary Fig. S2). Subsequently, because several datasets support multiple DE comparisons, we armed Fibromine with a controlled vocabulary of terms so as to be able to effectively codify and match all included phenotypes/DE comparisons across datasets. More specifically, based on respective literature and GEO retrieved nomenclature, all DEA comparisons were denoted by a three-part phrase, such as *A_vs_B* (or *A vs B*), where the first and third elements represent the two participating experimental conditions. Any dashes or underscores used in the comparisons pre-fix and suffix elements, *A* and *B*, separate sample from potential treatments and/or time-points. The capital letter *D* followed by one or multiple digits is used to denote the day of sample collection respective to the starting time-point of a given treatment.

Last, to enrich these data and support an unhampered by external parameters data exploration experience we have also included into Fibromine gene, transcript and protein level annotation retrieved from Ensembl¹⁶, UniProt/Swiss-Prot¹⁷ and STRING databases¹⁸, as well as gene ontology terms from Refs.^{19,20}, miRNA annotation downloaded from mirBase²¹ and miRNA target interaction predictions from miRDB²² (Fig. 1a). Human and mouse transcription factor annotation was obtained from Refs.^{23,24}, respectively. All operations were supported by in-house developed R/bash scripts and the use of GEOquery²⁵ and biomaRt²⁶ Bioconductor packages.

Conclusively, Fibromine is an ab initio created IPF-oriented database that interlinks a great volume of DEA results with rich out-sourced annotation. In addition, it is worth mentioning that Fibromine is the first lung fibrosis database that encompasses multi-omics and multi-species data, thus offering a great starting line for comparative meta-analyses.

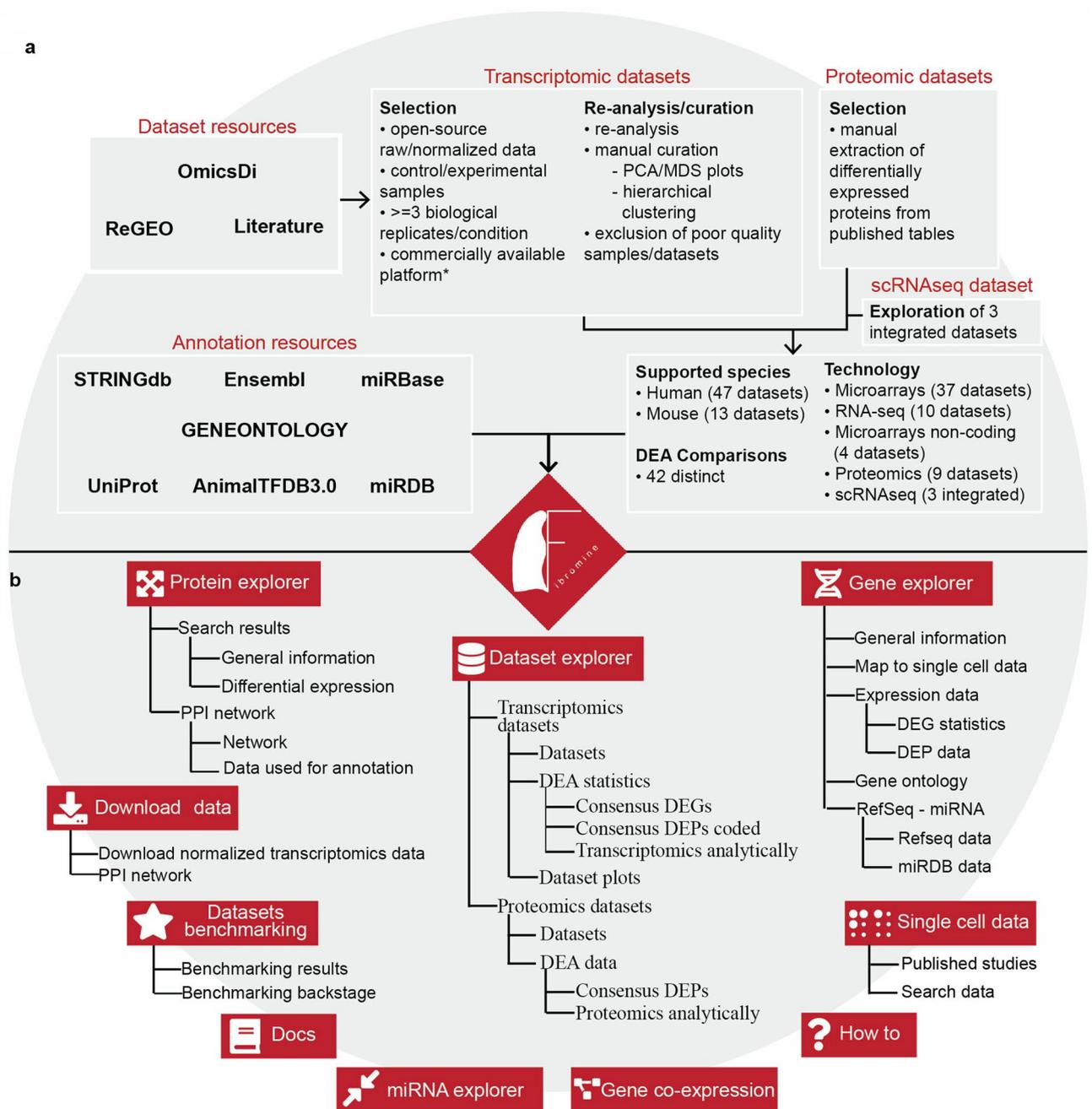


Figure 1. Fibromine database creation workflow and homonymous web server map. **(a)** Fibromine included datasets were discovered via literature, <https://www.omicsdi.org/> and <http://www.regeo.org/> scrutiny prior to specific criteria-based selection. Transcriptomic datasets were re-analysed and manually curated, while differentially expressed proteins were gleaned from respective publications. Differential expression was performed in a collection of three integrated scRNA-seq datasets. Annotation was sourced from various open-access resources. **(b)** Fibromine web interface is Shiny-powered. It is made out of three main explorers *Dataset*, *Gene* and *Protein explorer* and several smaller ones, while *Datasets benchmarking* tab provides a metrics-based system of datasets accreditation. *Single cell data* tab provides a list of relative published studies along with DEA results from three integrated datasets. The user can also map a gene directly to several single cell datasets via *Gene explorer*. *MDS*: multi-dimensional scaling, *PCA*: principal component analysis. Annotation sources: <https://string-db.org/>, <https://ensembl.org/>, <https://mirbase.org/>, <http://geneontology.org/>, <https://uniprot.org/>, <http://bioinfo.life.hust.edu.cn/AnimalTFDB/>, <http://mirdb.org/>. Figure was created using <https://www.google.com/slides/about/>.

Fibromine datasets benchmarking. As meta-analytic collections are prone to factors introducing biological/technical heterogeneity and IPF is a highly heterogeneous pathology, Fibromine's controlled vocabulary

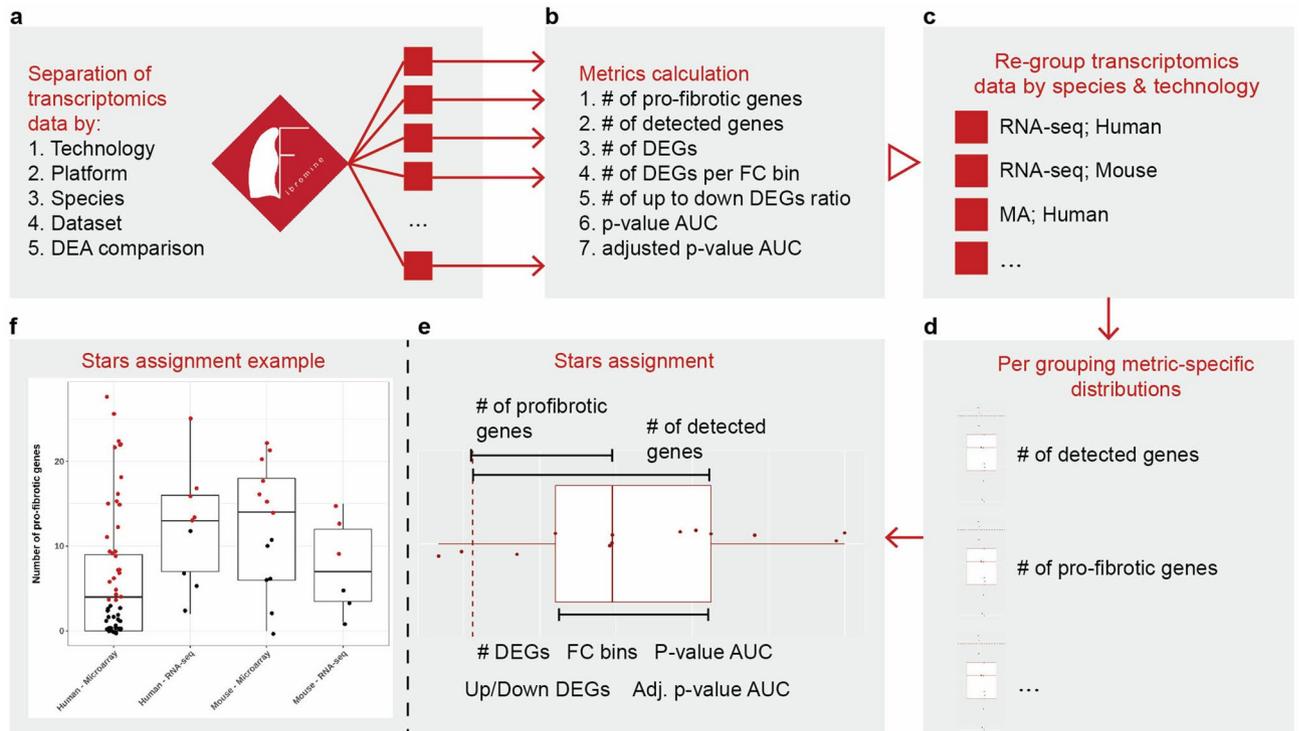


Figure 2. Fibromine transcriptomic datasets benchmarking workflow. In order to reveal the most homogeneous transcriptomic datasets of Fibromine, seven metrics were calculated (**b**) post to data separation (**a**). Subsequently, transcriptomic data were grouped per species and technology (**c**) and seven metric-specific distributions were shaped for each of the dataset groups (**d**). Finally, every dataset/DEA comparison was evaluated relative to the rest of its grouping and assigned a star if the calculated criterion value lied within a pre-specified interval of the respective distribution (**e**). Each dataset received 7 stars to the maximum and none to the minimum. Datasets with many stars are more closely related to each other than to the rest of the group. (**f**) A real example of stars assignment: red dots are datasets/DEA comparisons assigned a star for the *Number of pro-fibrotic genes* criterion. Datasets investigating exclusively non-protein coding genes were processed separately using the same workflow. Boxplots depict the interquartile range and median of the data; whiskers extend no longer than 1.5 times the length of the boxplot. *DEGs*: differentially expressed genes, *FC*: fold change, *AUC*: area under the curve. Figure was created using <https://ggplot2.tidyverse.org/> v3.5.5 and <https://www.google.com/slides/about/>.

may be up to a certain point capable of standardizing the DE comparisons included, but definitively cannot assess datasets underlying biological and technical differences that may affect any downstream integration attempt. For this reason, as well as the fact that most datasets' meta-data do not comprehensively describe crucial information that would assist in a variability reduction endeavour, instead of directly dealing with data variability we next set to offer a thorough means of datasets evaluation, easily interpretable by all users. For such a purpose, we designed a dataset assessment system, which we implemented on dataset/DEA groupings (Fig. 2a) in order to distinguish and accredit the most similar datasets within certain groups shaped using experimentally-based criteria. Our strategy is currently based on seven broadly used metrics: (i) the number of detected and (ii) differentially expressed genes, (iii) the representation of known pro-fibrotic genes among each dataset's Differentially Expressed (DE) genes, (iv) the number of genes with low $\{x, 1.2 < x < 2\}$, intermediate $\{x, 2 \leq x < 5\}$ or high $\{x, 5 \leq x\}$ absolute fold change, (v) the ratio of up- to down-regulated genes, as well as the area under their (vi) nominal and (vii) adjusted p-value distributions. Human pro-fibrotic genes already known to be implicated in pulmonary fibrosis were collected both from Ref.³ and from our lab's multi-year observations, while their homologue counterparts, as reported in Ensembl database, were used for murine datasets assessment.

More specifically, we separated all bulk transcriptomic data by technology, platform, species, dataset and DEA comparison and then calculated the aforementioned metrics (Fig. 2a,b). Afterwards, data were grouped per species/technology and metric-specific distributions were constructed (Fig. 2c,d). Last, datasets/DEA comparisons were accredited a star for every calculated value of a metric lying within a pre-specified distribution interval (7 stars at maximum) (Fig. 2e,f, Supplementary Fig. S3). All benchmarking results are recorded at *Dataset* and *Gene explorer* matrices, while an analytical description of the process along with intermediate products can be found at *Datasets benchmarking* tab of Fibromine. Datasets/DEA comparisons with more stars are most similar to each other relative to the rest of the group. Datasets investigating exclusively non-coding genes were processed separately using the same workflow. Finally, because our database addresses a wide variety of experimental conditions which are expected to affect benchmarking results when assessed simultaneously, we have repeated the

forementioned process exclusively for datasets belonging to important Fibromine comparisons (*IPF_vs_Ctrl* and *BleomD14_vs_Ctrl* on lung tissue) (Supplementary Fig. S4).

To conclude with, although not completely treating data variability, the realized datasets benchmarking strategy provides an extra layer of datasets comparative information, not readily available neither from raw, publicly available data, nor directly extractable from respective publications, based on which the most interesting/ “homogeneous” datasets can be picked for integration and/or mining.

Fibromine online interface: a web tool for IPF data integration and mining. A central pillar of the hereby described endeavour, is to facilitate data mining (mainly data summarization and aggregation) and comparative exploration of Fibromine data for the non-specialist user. For this reason, we have developed a Shiny-based online application that enables data access from three distinct but complementary points of observation: from a dataset (*Dataset explorer*), from a gene (*Gene explorer*; *miRNA explorer*) and from a protein level (*Protein explorer*). In addition to the aforementioned layers of information, *Gene co-expression* tab graphically displays significant modules of human and mouse lung-specific gene co-expression networks, while *Single cell data > Search data* tab presents in a tabular format DEA results for a great number of single cells.

Dataset explorer. To begin with, Fibromine’s *Dataset explorer* is divided into two tabs responsible for bulk transcriptomic (*Transcriptomic datasets*) and proteomic datasets (*Proteomic datasets*) (Fig. 1b), each one organised around a central interactive table. The user’s input is minimized to the selection of one (exploration) or multiple datasets of interest (integration) and the subsequent press of a button. Particularly, as far as transcriptomic datasets integration is concerned, Fibromine initially detects between datasets cDEGs (Fig. 3, Supplementary Fig. S5). Subsequently, it examines the available proteomic datasets matching the selected datasets experimental parameters for any consensus differentially expressed protein (cDEP) coded by any of the cDEGs. Addedly, in order to facilitate biological interpretation of the cDEG list, the user can perform over-representation analysis based on five gene-term libraries, exploiting the *Pathway analyses* tool provided. Concerning proteomic datasets integration, cDEPs are identified using the same pipeline of cDEGs determination. For more technical details about cDEGs and cDEPs, please, refer to “Methods” section of the paper.

Understanding the limitations and subjective nature of adopting strict statistical thresholds for consensus differentially expressed features (cDEFs) definition (“Methods” section), we have enabled the user to adjust the default fold change and hypothesis testing constraints used during transcriptomics data integration. This feature is not available for proteomic datasets, because the original publication thresholds were considered *as is* for the identification and isolation of Fibromine-included DE proteins. Moreover, the user can further influence the strictness of cDEFs report via the *Out of ... Datasets* column of *Consensus DEGs/Consensus DEPs* result tables, that controls the number of datasets out of the *n* user-selected ones where a feature’s expression has to be found consistently divergent, in order to be recognised as a cDEF (“Methods” section).

Last, as an extra means for transcriptomic datasets inspection, an interactive exploratory heatmap and volcano plot can be depicted upon demand at the *Dataset plots* tab of the explorer. There, the user can examine samples hierarchical clustering and create a volcano plot for each queried dataset. Selected markers of fibrosis are shown in the volcano plot providing a coarse-grained estimation of each dataset’s quality.

Gene explorer. Being able to acquire differential gene expression data across a great number of samples and datasets in a targeted fashion, is a catalytic step towards supporting wet lab findings and fuelling novel hypotheses formation. To painlessly retrieve the aforementioned information from Fibromine, we have implemented *Gene explorer* (Fig. 1b, Supplementary Fig. S6). Taking as input one or multiple genes of interest, *Gene explorer* displays the statistics (fold change, nominal and FDR adjusted p-value) of all their DE instances found in the database, along with information about their genomic position and biotype, related GO terms and RefSeq sequences. Moreover, as several miRNAs have been reported to be implicated in IPF pathology⁹, the display of miRDB-sourced mRNA-miRNA interactions relative to the queried protein-/miRNA-coding gene at the latest tab of the explorer can prove of extreme use. In parallel, the explorer displays data for any DEP coded by the queried genes. Last, in order to combine the better established methods of bulk DEA with the higher resolution of single cell data, *Gene explorer* maps each requested gene to single cell datasets of NU-Pulmonary cell browser in a species specific manner: human to Reyfman²⁷ and murine to Joshi-Watanabe²⁸ and Xie²⁹ datasets. More details on published scRNA-seq pulmonary fibrosis datasets can be found in the *Single cell data* tab of our web server, along with DEA results of Mayr et al. dataset¹³ in a tabular format. Please, refer to “Single cell data” section of the paper for more details.

miRNA explorer. Among the transcriptomic studies hosted by Fibromine, there are some having explicitly sequenced non-coding genes. Although these transcripts and their potential targets can be individually interrogated via *Gene explorer*, the latter does not simultaneously provide differential expression statistics for both coding and non-coding interactors. Thus, in order to accelerate the discovery of important regulatory interactions, we have created *miRNA explorer* (Supplementary Fig. S7). In more detail, we have integrated all *IPF_vs_Ctrl* lung non-coding array datasets and isolated the consensus differentially expressed miRNAs. Subsequently, we have performed likewise for their miRDB-sourced targets found at the bulk *IPF_vs_Ctrl* lung coding datasets. Both integrations were performed using the same pipeline applied to define cDEFs (“Methods” section) and $|FC| > 1.2$ and $FDR < 0.05$ values were used as differential expression thresholds. As a result, the user can mine the most important regulatory pathways taking place during lung fibrosis by selecting any member of the deregulated miRNA list and automatically obtain all mRNA targets characterized by an opposite direction of expression.

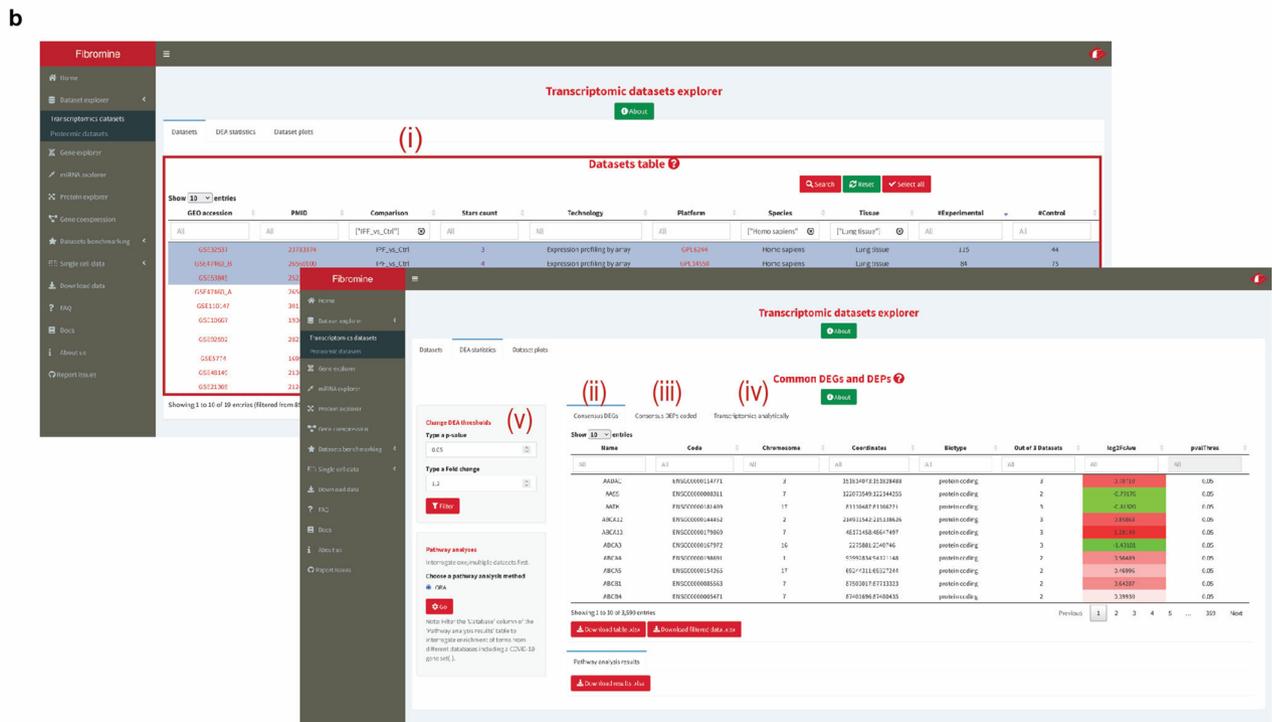
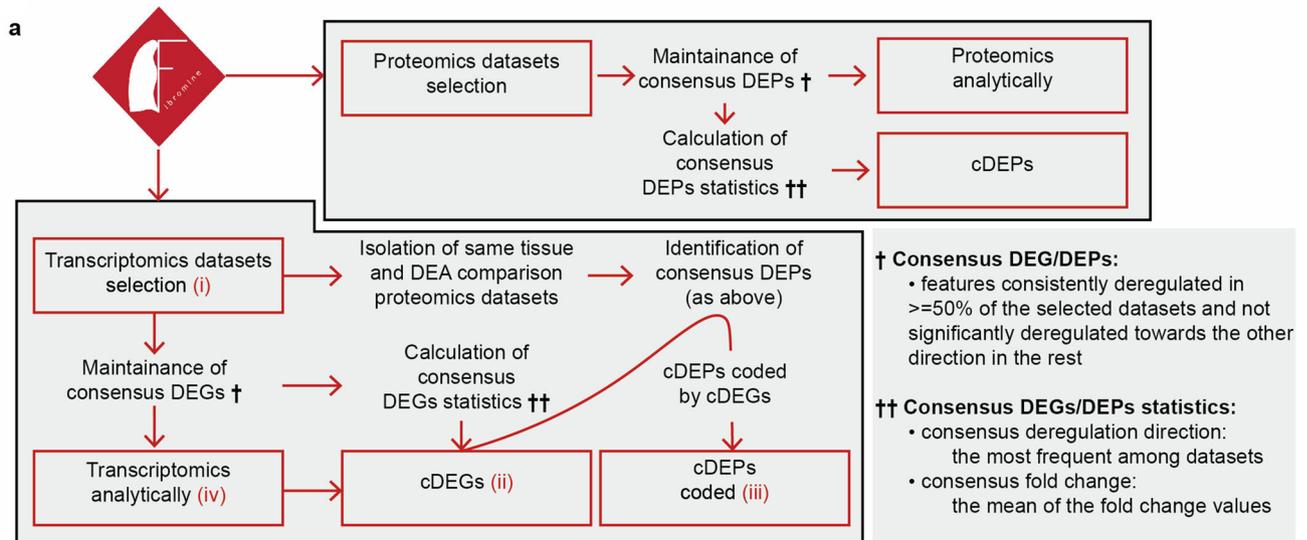


Figure 3. Same species datasets integration workflow. **(a)** Dataset explorer same species transcriptomic/proteomic datasets integration workflow. Boxes of red correspond to front-end elements, while the rest account for back-end processes. Latin number elements are presented in **(b)**. **(b)** Front end steps of transcriptomic datasets integration through *Dataset explorer*: (i) datasets selection, (ii) consensus DEGs table, (iii) consensus DEPs coded by any of the consensus DEGs and sharing the same direction of deregulation, (iv) analytical statistics table. Differential expression default parameters can be changed via a dedicated tuning panel (v). *lof2FcAve* column of (ii) holds the consensus direction of deregulation for each of the reported genes in a color coded fashion. *Out of... Datasets* column of the same table enables further tuning of consensus features identification procedure. *cDEGs*: consensus differentially expressed genes, *cDEPs*: consensus differentially expressed proteins. Figure was created using <https://www.google.com/slides/about/>.

Protein explorer. Although the IPF related proteomic datasets are significantly fewer than the respective transcriptomic ones¹⁴, the proteomic milieu may hold key leads for deciphering IPF progression. For this reason, we have equipped Fibromine with *Protein explorer* (Fig. 1b), a proteomic datasets accessor and protein–protein interaction network creation tool. Requiring as input only a protein’s coding gene name, the explorer initially presents some basic annotation for the queried protein, alongside the respective differential expression data, if any. Subsequently, aiming to inter-connect transcriptomic and proteomic expression data, the explorer hosts a

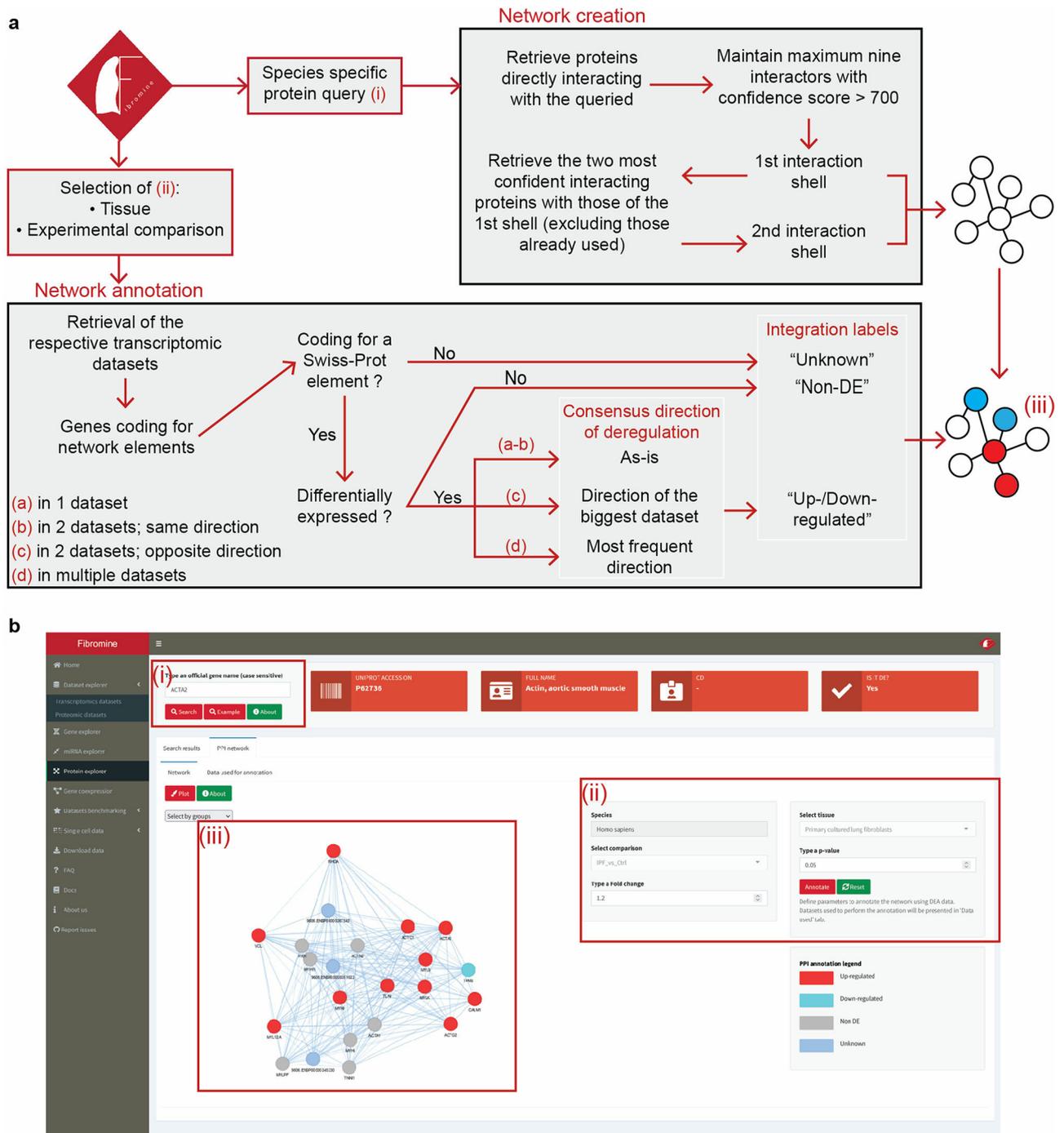


Figure 4. Protein–protein interaction network creation and condition-specific annotation workflow. **(a)** *Protein explorer* back-end process for the creation and annotation of protein–protein interaction networks. Data on protein relationship and interactions confidence are retrieved from Fibromine included UniProt data. Boxed and Latin numbers in red correspond to front-end elements presented in **(b)**. **(b)** The pipeline on the front-end of *Protein explorer*: (i) specific protein query, (ii) selection of experimental parameters to consider, (iii) annotated network. Figure was created using <https://www.google.com/slides/about/>.

condition-specific PPI network creation tool under the homonymous tab (Fig. 4). With the push of a button, the user can in real-time create a two-shell, high-confidence interactive network revolving around the queried protein and then automatically annotate it upon demand, based on the expression pattern of its nodes in any Fibromine-supported transcriptomic dataset/experimental comparison.

As far as network annotation is concerned, DE statistics for all genes coding for the UniProt/Swiss-Prot members of the network are initially isolated from the datasets corresponding to the experimental parameters selected by the user. Then, differential expression data are summarized for each of the genes into integration

labels holding the consensus direction of deregulation: *Non-DE*, *Up-regulated* and *Down-regulated*. Genes coding for TrEMBL proteins are assigned the *Unknown* label as Fibromine does not currently support TrEMBL entries. The aforementioned labels are used to color-code network nodes and thus transfer all available transcriptomics information into the protein level. Analytical statistics of the data used to annotate the network can be inspected at the *Data used* tab of the explorer, while it is worth mentioning that the default DEA thresholds are easily tunable to meet all users' expectations.

Gene co-expression. Gene co-expression analysis has been used successfully in the past for regulatory targets prediction in IPF⁹. Motivated by those previous attempts and in order to complement Fibromine's PPI network creation tool, we have designed *Gene co-expression* tab to host a human (*IPF_vs_Ctrl*) and a mouse (*BleomycinD14_vs_Ctrl*) lung-specific gene co-expression network (GCN) (Supplementary Fig. S8). Through this tool, with only the selection of any, potentially, disease driving gene, as pinpointed via Weighted Gene Co-expression Network Analysis (WGCNA), the user can plot interactive GCNs spanning any of the top fibrotic phenotype-related modules. Each network is built on the members of the module where the selected feature belongs. Network nodes represent genes with a module membership (MM) and gene significance (GS) above the 60th percentile of their respective module, while edges represent high confidence gene relationships, with a Topological Overlap Measure (TOM) above the 3rd quartile of the respective module's distribution. The queried gene is marked in red and it is the only node allowed not to have any edge. Last, both MM and GS thresholds can be tuned via Fibromine to dynamically change the strictness of node selection.

Single cell data. As mapping of genes at the single cell level through the *Gene expression* tab is limited to visual inspection of feature plots, we designed the *Single cell data* tab in order to provide detailed DEA statistical data from one of the biggest datasets regarding lung fibrosis¹³ (233,638 cells from 10 pulmonary diseases and control donor samples). More specifically, through *Search data* sub-tab the user can access DEA data of the top 2000 most variable genes, examined for differentiating expression levels between each pathology and the control cells. These data are expected to increase resolution of the bulk sequencing data presented at the *Gene explorer* tab, as well as to generalize DEA findings to the greater niche of pulmonary pathologies.

Overall, Fibromine web server constitutes a data mining and integration portal suitable for all users irrespective their computational background. As far we are able to know, its highly-automated exploration features render Fibromine, the first IPF-revolving toolkit able to integrate data across species and experimental designs, map them to the single cell level in a visual and tabular format and to offer a quick and tangible way of transcriptomics data projection to an environment of proteomics intercommunication. Last but not least, two other prominent functionalities of Fibromine is mining and presentation of the most biologically promising lung fibrosis-specific miRNA-mRNA interactions and gene co-expression networks.

Usage of retrieved individual gene expression patterns for novel hypothesis formation. Fibromine's *Gene explorer* enables interrogation of single/multiple gene expression motifs (consistently recurring DEA results across same/similar experimental conditions), a very useful feature to validate wet laboratory findings and quickly deepen novel hypotheses formation. For example, following leads from cancer and inflammatory diseases, our laboratory has used a primary version of Fibromine's *Gene explorer* to discover that based on the three larger IPF_vs_Ctrl bulk lung datasets, *MAP3K8* is down-regulated in IPF patients' lungs³⁰. In addition, inspecting *MAP3K8* in higher resolution data via the explorer's *Map to single cell data* feature instantly revealed an "enrichment" of gene's expression in monocytic-lineage cells. These findings were further validated in the bleomycin-induced animal model of pulmonary fibrosis, where genetic deletion of *MAP3K8* was shown to exacerbate the modeled disease³⁰. All results impelling us to wet lab examination of *MAP3K8* expression during lung fibrosis can be replicated using the *Example* button of *Gene explorer* and then filtering the DEG statistics table for lung tissue and IPF_vs_Ctrl comparisons (Supplementary Fig. S6).

Identification of "bona fide" differentially expressed genes in human and mice fibrotic lungs. Another very useful function of Fibromine's web server is its ability to encapsulate differential gene expression data from a great number of samples and datasets sharing similar experimental parameters. This feature is particularly important as it extends over the limited sample size of single datasets and yields results reported by several research efforts. *IPF_vs_Ctrl* and *BleomycinD14_vs_Ctrl* human and mouse lungs respectively, constitute the two most numerous of the 42 unique differential expression comparisons available through Fibromine. To retrieve a list of potential fibrosis drivers for each of these experimental settings, we have initially pinpointed through *Datasets benchmarking* tab (Fig. 1b) those *IPF_vs_Ctrl* human lung datasets with at least 4 stars (Fig. 5a) and the three top accredited *BleomycinD14_vs_Ctrl* lung murine datasets (Fig. 5b). Then, we integrated via *Dataset explorer* the human datasets retrieving a list of 2182 human cDEGs, prior to mouse datasets separate integration which displayed a longer collection of 3863 cDEGs. Both lists can be exported through Fibromine online application using the default thresholds ($|FC| > 1.2$ and $p\text{-value} < 0.05$) and requiring as input only the selection of the respective datasets from *Transcriptomics datasets* tab main interactive table. For more details on back-end processes, please, refer to respective sections of the publication.

Across-species datasets integration. Bleomycin-induced pulmonary fibrosis in mouse is a well-established and broadly used model of pulmonary fibrosis³¹ and thus it is expected to bear common ground with the human disease. The aforementioned lists of consensus differentially expressed genomic features provide the opportunity for an inspection of model's fitness to simulate human pathology, but such a comparative undertaking surely requires extra manual processing. To avoid such a limitation, Fibromine enables automatic between

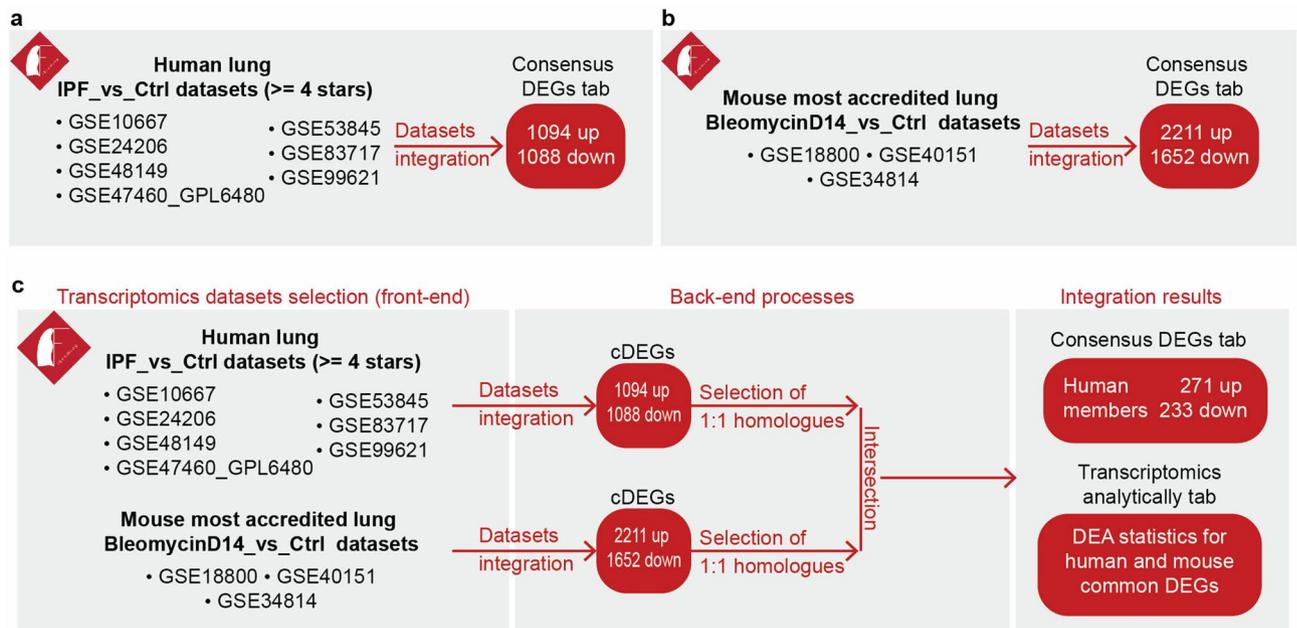


Figure 5. *IPF_vs_Ctrl* lung and *BleomycinD14_vs_Ctrl* datasets integration. **(a)** Integration of the *IPF_vs_Ctrl* lung datasets with at least 4 accreditation stars. **(b)** Integration of the top 3 starred *BleomycinD14_vs_Ctrl* datasets. **(c)** Across-species integration of the **(a,b)** datasets. Selection of the top accredited human and mouse transcriptomics datasets for integration leads to the identification of 504 consensus DEGs. All intermediate and integration output lists can be retrieved through Fibromine. Integration was performed using default differential expression thresholds. *cDEGs*: consensus differentially expressed genes, *DEA*: differential expression analysis. Figure was created using <https://www.google.com/slides/about/>.

species integration of datasets upon their simultaneous selection through the explorer's interactive table (Supplementary Fig. S5). As can be seen from Fig. 5c, approximately one quarter of the initial 2182 human cDEGs were found to have a consistently deregulated mouse homologue, with their 1:1 across species homology being an extra factor advocating for a more probable cross-species direct relationship at the molecular level. Cross-species cDEGs can be recreated through Fibromine via integration of the aforementioned datasets.

Discussion

As more and more omics datasets addressing divergence from steady state emerge and accrue to the existing ones, a big bet for the contemporary biomedical sciences is to be able to fruitfully mine latent information from these already available wealthy data sources in order to guide, accelerate and validate wet laboratory research. This strategy can prove extremely useful especially for rare diseases such as IPF, which logically attract less attention than more prevalent ones. Moreover, biomedical sciences have nowadays entered an era of single cell centred research, enabling pathologies exploration at an unprecedented level of resolution. Nevertheless, pulmonary fibrosis research still lacks a central resource that would help scientists to exploit the vast legacy of bulk omics and single cell datasets and their inherent characteristic such as a smaller drop-out effect³². Furthermore, as deconvolution methods have proven to be affected by data pre-processing³³ steps, a comprehensive collection of consistently handled bulk data could also facilitate the verification of single cell level observations by providing the bulk sequencing material necessary for such analyses.

Taking all the above into consideration, we have developed Fibromine, a database of IPF omics datasets of both transcriptomic and proteomic nature, spanning human and murine research deliverables of various designs. For this purpose we initially manually selected, consistently re-analyzed and carefully curated a great number of bulk transcriptomic and proteomic datasets which consist the backbone of our database. The latter, also includes wealthy third-party annotation and dataset meta-data yielded from literature. Smooth access to Fibromine is ensured by a Shiny-powered web server that simultaneously examines data from multiple complementary perspectives: from a dataset, a single gene/protein, a gene co-expression and protein-protein interaction point-of-view. Crucially, bulk sequencing resolution is increased via provision of in-house single cell level data organized in a tabular format and dynamical connection to external single cell data resources for visual examination. Last, for those users who wish to perform their own downstream analysis, our server supports download of all included data and integration/mining results.

One of the main assets of the hereby presented project is the freedom it provides to the user to fine-tune crucial exploration and integration parameters. To begin with, the user can experiment changing the default DEA thresholds which in turn affect, amongst others, the features of dataset integration and PPI network annotation. In addition, by consulting the available datasets meta-data, the gene/protein annotation included, and more importantly our transcriptomic datasets accreditation system, the user has a quick and tangible means of shaping a well calculated decision regarding which data are contextually best to investigate. Most importantly, thanks to

our datasets benchmarking system, the user has a tangible means of comparing transcriptomic datasets based on their otherwise not readily accessible characteristics.

Apart from the well-established data analysis methods used, the operation of Fibromine website is subject to certain assumptions. First, instead on inspecting corrected p-values for cDEG identification, Fibromine uses their nominal counterparts in order to avoid exclusion of borderline significant features, which may be however reported as deregulated using the looser uncorrected threshold in multiple of the user integrated datasets, thus indicating a latent biologically significant feature. In parallel, as multiple testing bias indisputably needs to be taken into consideration, corrected per-dataset statistics for the features reported are presented in the *Transcriptomics* or *Proteomics analytically* tabs. We strongly encourage consultation of the latter prior to any wet or dry laboratory downstream procedure. Furthermore, although very informative, some of the metrics used during datasets benchmarking are based on some general assumptions, such as that the murine homologues of human pro-fibrotic genes exert the same effect in the mouse IPF model and thus, may require future fine-tuning. As far as PPI networks annotation is concerned, the decision of using transcriptomic instead of proteomic data was determined by: (i) the tremendously smaller number of supported DE proteins (693 unique ones) compared to the number of described DE genes (17,153 unique DE protein coding genes; $|FC| \geq 1.2$ and FDR corrected p-value < 0.05) and (ii) the currently limited number of included proteomic data experimental designs compared to those of transcriptomics nature. Therefore, use of the former instead of the latter would have surely limited PPI annotation options and would have led to the characterization of an excess of nodes as *Non-DE*, thus rendering the tool rather impractical.

Further work needs to be done for Fibromine's active maintenance and expansion. Despite the careful literature inspection, some of the latest transcriptomic datasets may not have been included into our database and some minor DE comparisons supported by the currently included datasets have not yet been examined. On top of that, our database currently addresses an important yet incomplete portion of proteomic data (*IPF_vs_Ctrl* human datasets) which will also be expanded to include invaluable datasets such as Ref.³⁴. Last but not least, enrichment of Fibromine with datasets from other omics technologies such as metabolomics, lipidomics and predominantly a greater number of single cell datasets are some of our top priorities.

Conclusively, this endeavour set out to implement from scratch a centralized web resource for the acceleration of IPF research. Through Fibromine, both computational and most importantly non-computational background supported biomedical scientists can quickly and effortlessly obtain, integrate and compare information regarding DE events during a wide variety of IPF-related conditions. Hopefully, Fibromine will prove itself the driving force for novel hypothesis formation and new biomarkers discovery.

Methods

Datasets re-analysis and curation. Microarray datasets re-analysis was performed using ad hoc developed pipelines based on well-established tools and methods, while metaseqR2 Bioconductor package³⁵ was used for the re-analysis of RNA-seq data.

For microarrays data processing, limma³⁶, oligo³⁷ and beadarray³⁸ Bioconductor packages consisted the analysis workhorses while arrayQualityMetrics package was utilized for quality control purposes³⁹. More specifically, probe intensities were within-technology consistently background corrected, within- and across-technologies uniformly normalized using quantile normalization and then summarized to the gene level using a weighted average. Control sequences and probes mapping to multiple HGNC gene symbols were removed from further analysis. DEA was conducted for all datasets using the limma moderated t-test statistics method.

RNA-seq fastq files were mapped to GRCh38.p13 and GRCm38.p6 genomes using a two steps alignment pipeline that exploits HISAT2⁴⁰ and Bowtie2⁴¹ aligners. Initially, HISAT2 was used to map reads in a splice aware fashion, while those failing to align were delivered to Bowtie2 for a second, more sensitive alignment round. Further processing was conducted using the Bioconductor package metaseqR2³⁵: raw counts files were quantified at the gene level, normalized with EDASeq⁴², filtered using default parameters and statistically analysed for DE using PANDORA algorithm. The latter combined the results of the well-known and broadly used DESeq⁴³, DESeq2⁴⁴, edgeR⁴⁵, limma-voom³⁶ and ABSSeq⁴⁶ DEA methods, leading to more precise lists of differentially expressed genomic features. More precisely, PANDORA combines individual DEA algorithms p-values in a weighted manner, with weights shaped according to each method's performance during real-data based simulations. As a result, among others, PANDORA achieves a better precision-recall trade-off and reduces the effects of gene length in downstream analyses, such as pathway analysis.

For microarray and RNA-seq datasets, sample outliers were identified prior to removal using PCA and MDS plots, respectively, as well as samples hierarchical clustering of log₂-scaled normalized values of the thousand top DE genes. The later were defined using the thresholds of $|FC| > 1.2$ and a significant p-value at an FDR threshold of 5%.

Proteomics data were retrieved directly from published tables and DE thresholds were kept *as is* in each original publication. An exception was made for a single dataset⁴⁷ due to the very limited number of DE proteins returned otherwise. Proteins quantified in mixtures/same aptamer were removed.

Identification of consensus differentially expressed features. During same species datasets DEA results integration, a feature (either transcript or protein) is called consensus differentially expressed if it has been found consistently deregulated (in the same direction of deregulation) in at least half of the user-selected datasets and not significantly deregulated towards the opposite direction in any of the rest; *Consensus DEGs/Consensus DEPs* tabs. According to those, we report as consensus direction of deregulation, the one most frequently encountered across them, while as consensus fold change the mean of their fold change values. When integrating datasets from multiple species, a feature is defined as consensus DE if it has 1:1 human:mouse homology (based

on Ensembl data) and has been found consistently deregulated in at least half of each species' selected datasets. Datasets examining explicitly non-coding genes are not considered for across species integration. Consensus fold change is calculated as in the case of same species data. For reasons of clarity, only the human component of across species datasets integration is summarized in *Datasets explorer's Consensus DEG* and *Consensus DEPs coded* tabs. In both the above cases, DEA thresholds are $|FC| > 1.2$ and $p\text{-value} < 0.05$ for genes and those applied during original analysis for proteins. For more on hypothesis testing thresholds selection, please, refer to “Discussion” section of the paper. All thresholds for DE genes identification are user-tunable from within Fibromine.

Pathway analysis. Fibromine's *Dataset explorer* supports pathway analysis of cDEGs. Specifically, enrichR package is used to connect with the Enrichr database and perform over-representation analysis for up and down regulated genomic features separately, based on five libraries: KEGG, BP GO, MF GO, BP GO and one of COVID-19-related associations. For more on current use and implementation of the aforementioned gene set libraries please refer to the latest Enrichr publication⁴⁸.

Heatmap and volcano plot creation. Dataset-specific Fibromine's interactive heatmaps and volcano plots are crafted using heatmaply⁴⁹ and plotly R package⁵⁰, respectively. Clustering of each dataset's samples is performed on z-score scaled log₂-transformed normalized expression values of the top one thousand DE genes using Euclidean distance and complete linkage for clusters comparison. Volcano plots specify deregulated genes based on the thresholds of $|FC| > 1.2$ and $p\text{-value} < 0.05$. A significant, yet non-comprehensive list of genes known to be implicated in pulmonary fibrosis are indicated in every volcano plot.

Condition-specific protein–protein interaction networks creation. PPI networks are shaped according to Fibromine-incorporated STRING database¹⁸ data. More specifically, each network consists of two interaction shells, with the first including maximum nine, high confidence (interaction score > 700) proteins interacting with the queried one. For the determination of second shell elements, the two most confident interactors for each of the first shell proteins are selected. Proteins already-considered for the first shell creation are not considered for the second one. All networks have a DrL layout with protein interaction scores used as edge weights.

For the condition-specific annotation of networks, genomic features coding for the network elements are identified and their differential gene expression statistics are recovered from the datasets corresponding to user-selected experimental parameters. From the aforementioned genes only those corresponding to UniProt/Swiss-Prot members are retained and for the cases of 1:many UniProt/Swiss-Prot:gene entries the gene featured by UniProt as primary is utilized. Afterwards, each network node is assigned a label of a consensus direction of deregulation: “Unknown” if the gene product is not a Swiss-Prot member, “Non DE” if the corresponding gene is not DE and “Upregulated” or “Downregulated” for the deregulation cases. The latter two labels are shaped according to the following rules: if a gene is found DE in a single dataset or consistently deregulated in a couple of datasets, its direction of deregulation is kept *as is*; if a gene is found DE in a couple of datasets with an inconsistent direction of deregulation between the two, the biggest dataset's direction of deregulation is maintained; if a gene is found DE in multiple datasets, then the most frequent direction is utilized. By default, DE genes are defined as those having an absolute fold change bigger than 1.2 and a p-value smaller than 0.05. Multiple testing corrected statistics are presented at the *Data used* tab of the explorer.

Gene co-expression networks creation. For the human network, *IPF_vs_Ctrl* lung tissue datasets with more than four stars assigned were selected (GSE10667, GSE24206, GSE48149, GSE47460_GPL6480, GSE53845, GSE83717, GSE99621), while the three most accredited *BleomycinD14_vs_Ctrl* were chosen for the mouse one (GSE18800, GSE40151, GSE34814). Normalized, z-transformed gene expression values were applied to determine the scale-free co-expression modules via WGCNA⁵¹. Biweight midcorrelation, a more robust alternative of Pearson coefficient, was used to create a signed network, while signed Topological Overlap Measure (TOM) was calculated by adjacency transformation to decrease noise. Network modules were identified using (1-TOM) as a distance metric and module eigengene (ME) was determined using default parameters. Closely clustered modules (0.25 and 0.40 distance determined by hierarchical clustering for human and mouse, respectively) were merged. MEs were correlated to the trait of interest (fibrotic/non-fibrotic tissue) and those with a significant Pearson correlation coefficient were maintained ($|r| > 0.6$; $p\text{-value} < 0.05$). Potential phenotype-drivers were pinpointed by intra-modular analysis having a statistically significant module membership and gene significance value ($p\text{-value} < 0.05$). Pearson co-efficient for the latter two metrics can be adjusted via Fibromine server for a more thorough and objective features selection.

For the creation of network visualisations, all features of the selected gene's module with module membership and gene significance above the (default) 60th percentile are selected. Network edges represent correlations in the 3rd quartile of the pre-calculated TOM and with the exception of the queried gene, all genes having a zero degree of connectivity are dropped prior to minimum spanning tree calculation and network visualization. Network layout is automatically chosen and the thresholds of module membership and gene significance are user tunable.

Single cell data analysis. Seurat package v.4 was used for the analysis of single cell data found [here](#). In detail, “empty” cells, as annotated during original analysis, were removed prior to identification of the top 2000 most variable genes using the *vst* method (FindVariableFeatures). Finally, Wilcoxon rank sum test was used for DEA of the above mentioned genes using default parameters (FindMarkers).

Data availability

The data that support this study can be accessed freely via <http://www.fibromine.com/Fibromine>, a Shiny-based web tool.

Code availability

The code for the creation and operation of Fibromine web server is available via <https://github.com/dfanidis/Fibromine>.

Received: 6 August 2021; Accepted: 21 October 2021

Published online: 05 November 2021

References

- Antoniou, K. M. *et al.* Precision medicine in idiopathic pulmonary fibrosis therapy: From translational research to patient-centered care. *Curr. Opin. Pharmacol.* **57**, 71–80. <https://doi.org/10.1016/j.coph.2020.12.007> (2021).
- Somogyi, V. *et al.* The therapy of idiopathic pulmonary fibrosis: What is next? *Eur. Respir. Rev.* **28**, 153 (2019).
- Vukmirovic, M. & Kaminski, N. Impact of transcriptomics on our understanding of pulmonary fibrosis. *Front. Med.* **5**, 87 (2018).
- Raghu, G., Chen, S.-Y., Hou, Q., Yeh, W.-S. & Collard, H. R. Incidence and prevalence of idiopathic pulmonary fibrosis in US adults 18–64 years old. *Eur. Respir. J.* **48**, 179. <https://doi.org/10.1183/13993003.01653-2015> (2016).
- Maher, T. M. & Strek, M. E. Antifibrotic therapy for idiopathic pulmonary fibrosis: Time to treat. *Respir. Res.* **20**, 205. <https://doi.org/10.1186/s12931-019-1161-4> (2019).
- Herazo-Maya, J. D. *et al.* Peripheral blood mononuclear cell gene expression profiles predict poor outcome in idiopathic pulmonary fibrosis. *Sci. Transl. Med.* **5**, 205. <https://doi.org/10.1126/scitranslmed.3005964> (2013).
- Kim, S. *et al.* Integrative phenotyping framework (iPF): Integrative clustering of multiple omics data identifies novel lung disease subphenotypes. *BMC Genomics* **16**, 924. <https://doi.org/10.1186/s12864-015-2170-4> (2015).
- Li, D., Liu, Y. & Wang, B. Identification of transcriptomic markers for developing idiopathic pulmonary fibrosis: An integrative analysis of gene expression profiles. *Int. J. Clin. Exp. Pathol.* **13**, 1698–1706 (2020).
- McDonough, J. E. *et al.* Gene correlation network analysis to identify regulatory factors in idiopathic pulmonary fibrosis. *Thorax* **74**, 132. <https://doi.org/10.1136/thoraxjnl-2018-211929> (2019).
- Wang, Z., Zhu, J., Chen, F. & Ma, L. Weighted gene coexpression network analysis identifies key genes and pathways associated with idiopathic pulmonary fibrosis. *Med. Sci. Monit.* **25**, 4285–4304. <https://doi.org/10.12659/MSM.916828> (2019).
- Villaseñor-Altamirano, A. B. *et al.* PulmonDB: A curated lung disease gene expression database. *Sci. Rep.* **10**, 514. <https://doi.org/10.1038/s41598-019-56339-5> (2020).
- Chen, G. *et al.* Restructured GEO: Restructuring gene expression omnibus metadata for genome dynamics analysis. *Database* **2019**, 1. <https://doi.org/10.1093/database/bay145> (2019).
- Mayr, C. H. *et al.* Integrative analysis of cell state changes in lung fibrosis with peripheral protein biomarkers. *EMBO Mol. Med.* **13**, e12871. <https://doi.org/10.15252/emmm.202012871> (2021).
- Norman, K. C., Moore, B. B., Arnold, K. B. & O'Dwyer, D. N. Proteomics: Clinical and research applications in respiratory diseases. *Respirology* **23**, 993–1003. <https://doi.org/10.1111/resp.13383> (2018).
- Khan, T., Dasgupta, S., Ghosh, N. & Chaudhury, K. Proteomics in idiopathic pulmonary fibrosis: The quest for biomarkers. *Mol. Omics* **17**, 43–58. <https://doi.org/10.1039/D0MO00108B> (2021).
- Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res.* **48**, D682–D688. <https://doi.org/10.1093/nar/gkz966> (2020).
- UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515. <https://doi.org/10.1093/nar/gky1049> (2018).
- Szklarczyk, D. *et al.* STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613. <https://doi.org/10.1093/nar/gky1131> (2019).
- Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29. <https://doi.org/10.1038/75556> (2000).
- Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338. <https://doi.org/10.1093/nar/gky1055> (2019).
- Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: From microRNA sequences to function. *Nucleic Acids Res.* **47**, D155–D162. <https://doi.org/10.1093/nar/gky1141> (2019).
- Chen, Y. & Wang, X. miRDB: An online database for prediction of functional microRNA targets. *Nucleic Acids Res.* **48**, D127–D131. <https://doi.org/10.1093/nar/gkz757> (2020).
- Lambert, S. A. *et al.* The human transcription factors. *Cell* **172**, 650–665. <https://doi.org/10.1016/j.cell.2018.01.029> (2018).
- Hu, H. *et al.* AnimalTFDB 3.0: A comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* **47**, D33–D38. <https://doi.org/10.1093/nar/gky822> (2019).
- Davis, S. & Meltzer, P. S. GEOquery: A bridge between the gene expression omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847. <https://doi.org/10.1093/bioinformatics/btm254> (2007).
- Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191. <https://doi.org/10.1038/nprot.2009.97> (2009).
- Reyfman, P. A. *et al.* Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 1517–1536. <https://doi.org/10.1164/rccm.201712-2410OC> (2018).
- Joshi, N. *et al.* A spatially restricted fibrotic niche in pulmonary fibrosis is sustained by M-CSF/M-CSFR signaling in monocyte-derived alveolar macrophages. *Eur. Respir. J.* **55**, 1900646. <https://doi.org/10.1183/13993003.00646-2019> (2019).
- Xie, T. *et al.* Single-cell deconvolution of fibroblast heterogeneity in mouse pulmonary fibrosis. *Cell Rep.* **22**, 3625–3640. <https://doi.org/10.1016/j.celrep.2018.03.010> (2018).
- Zannikou, M. *et al.* MAP3K8 regulates Cox-2-mediated prostaglandin E2 production in the lung and suppresses pulmonary inflammation and fibrosis. *J. Immunol.* **206**, 607–620. <https://doi.org/10.4049/jimmunol.2000862> (2021).
- Barbayanni, I., Ninou, I., Tzouveleki, A. & Aidinis, V. Bleomycin revisited: A direct comparison of the intratracheal micro-spraying and the oropharyngeal aspiration routes of bleomycin administration in mice. *Front. Med.* **5**, 269 (2018).
- Baruzzo, G., Patuzzi, I. & Di Camillo, B. SPARSim single cell: A count data simulator for scRNA-seq data. *Bioinformatics* **36**, 1468–1475. <https://doi.org/10.1093/bioinformatics/btz752> (2020).
- Avila Cobos, F., Vandesompele, J., Mestdagh, P. & De Preter, K. Computational deconvolution of transcriptomics data from mixed cell populations. *Bioinformatics* **34**, 1969–1979. <https://doi.org/10.1093/bioinformatics/bty019> (2018).
- Schiller, H. B. *et al.* Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair. *Mol. Syst. Biol.* **11**, 819. <https://doi.org/10.15252/msb.20156123> (2015).
- Fanidis, D. & Moulos, P. Integrative, normalization-insusceptible statistical analysis of RNA-Seq data, with improved differential expression and unbiased downstream functional analysis. *Brief. Bioinform.* <https://doi.org/10.1093/bib/bbaa156> (2020).

36. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47. <https://doi.org/10.1093/nar/gkv007> (2015).
37. Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363–2367. <https://doi.org/10.1093/bioinformatics/btq431> (2010).
38. Dunning, M. J., Smith, M. L., Ritchie, M. E. & Tavaré, S. Beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics* **23**, 2183–2184. <https://doi.org/10.1093/bioinformatics/btm311> (2007).
39. Kauffmann, A., Gentleman, R. & Huber, W. ArrayQualityMetrics—A bioconductor package for quality assessment of microarray data. *Bioinformatics* **25**, 415–416. <https://doi.org/10.1093/bioinformatics/btn647> (2009).
40. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915. <https://doi.org/10.1038/s41587-019-0201-4> (2019).
41. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. <https://doi.org/10.1038/nmeth.1923> (2012).
42. Risso, D., Schwartz, K., Sherlock, G. & Dudoit, S. GC-content normalization for RNA-seq data. *BMC Bioinform.* **12**, 480–480. <https://doi.org/10.1186/1471-2105-12-480> (2011).
43. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106. <https://doi.org/10.1186/gb-2010-11-10-r106> (2010).
44. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21. <https://doi.org/10.1186/s13059-014-0550-8> (2014).
45. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297. <https://doi.org/10.1093/nar/gks042> (2012).
46. Yang, W., Rosenstiel, P. C. & Schulenburg, H. ABSSeq: A new RNA-Seq analysis method based on modelling absolute expression differences. *BMC Genomics* **17**, 1–14. <https://doi.org/10.1186/s12864-016-2848-2> (2016).
47. Saraswat, M. *et al.* Label-free plasma proteomics identifies haptoglobin-related protein as candidate marker of idiopathic pulmonary fibrosis and dysregulation of complement and oxidative pathways. *Sci. Rep.* **10**, 1–11. <https://doi.org/10.1038/s41598-020-64759-x> (2020).
48. Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* **14**, 128. <https://doi.org/10.1186/1471-2105-14-128> (2013).
49. Galili, T., O'Callaghan, A., Sidi, J. & Sievert, C. heatmaply: An R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics* **34**, 1600–1602. <https://doi.org/10.1093/bioinformatics/btx657> (2018).
50. Sievert, C. *Interactive Web-Based Data Visualization with R, Plotly, and Shiny* (Chapman and Hall/CRC, 2020).
51. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* <https://doi.org/10.1186/1471-2105-9-559> (2008).

Acknowledgements

The authors would like to thank all members of Vassilis Aidinis laboratory: Christiana Magkrioti, Ph.D.; Ilianna Barbayianni, Ph.D. candidate; Apostolos Galaris, Ph.D. candidate; Voula Kanellopoulou, Ph.D. candidate; Ellie Stylianaki, Ph.D. candidate; as well as the MSc students Kostas Ntatsoulis, Dimitris Nastos, Katerina Touloumi and BSc student Maria Tsiokou, for beta testing and constructive feedback. Special thanks to Evangelos Karatzas, Ph.D. for constructive suggestions for Fibromine's web server amelioration.

Author contributions

D.F., P.M. and V.A. designed the research. D.F. performed datasets re-analyses and curation, designed and implemented the Fibromine database and web server. D.F., P.M. and V.A. designed the datasets benchmarking strategy, while the first realised it. D.F. and V.A. wrote the manuscript. D.F. created the figures. All authors reviewed the manuscript.

Funding

This work was supported by European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call Research—Create—Innovate [Grant Number: T1EDK-0049 to V.A.]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-01069-w>.

Correspondence and requests for materials should be addressed to P.M. or V.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

An explainable machine learning-driven proposal of pulmonary fibrosis biomarkers

Μια εξηγήσιμη πρόταση βιοδεικτών πνευμονικής ίνωσης βασισμένη σε μηχανιστική μάθηση

Ασθένειες πνευμονικής ίνωσης βρίσκονται στο επίκεντρο της βιοϊατρικής έρευνας εξαιτίας τόσο του αυξανόμενου επιπολασμού τους, όσο και της συσχέτισης τους με τις λοιμώξεις SARS-CoV-2. Η έρευνα της ιδιοπαθούς πνευμονικής ίνωσης, της πιο θανάσιμης μεταξύ των διάμεσων πνευμονοπαθιών, έχει την ανάγκη για νέους βιοδείκτες και στόχους, μια διαδικασία που θα μπορούσε να επιταχυνθεί με τη χρήση τεχνικών μηχανιστικής μάθησης. Σε αυτή την μελέτη χρησιμοποιήσαμε τιμές Shapley για να εξηγήσουμε τις αποφάσεις που πάρθηκαν από ένα ensemble μοντέλο μάθησης εκπαιδευμένο να κατηγοριοποιεί δείγματα ως ινωτικά ή υγιή βάση τιμών μεταγραφής διαφορικά εκφρασμένων γονιδίων. Αυτή η διεργασία κατέληξε σε ένα πλήρες και ένα λακωνικό σύνολο γενετικών χαρακτηριστικών, ικανών να διαχωρίσουν φαινότυπους σε βαθμό τουλάχιστον αξιότιμο με αυτόν προηγουμένως δημοσιευμένων συνόλων βιοδεικτών. Ενδεικτικώς επετεύχθη μία μέγιστη αύξηση της τάξης του 6% στην ακρίβεια και 5% στον δείκτη συσχέτισης του Matthew's. Αξιολόγηση ενός επιπρόσθετου ανεξάρτητου συνόλου δεδομένων έδειξε πως το εδώ προτεινόμενο σύνολο χαρακτηριστικών έχει μεγαλύτερο δυναμικό γενίκευσης σε σχέση με τα υπόλοιπα. Εν τέλει, οι προτεινόμενες λίστες γονιδίων αναμένεται όχι μόνο να υπηρετήσουν ως νέα σύνολα διαγνωστικών δεικτών αλλά και ως μια δεξαμενή στόχων για μελλοντικές ερευνητικές πρωτοβουλίες.



Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research Article

An explainable machine learning-driven proposal of pulmonary fibrosis biomarkers



Dionysios Fanidis^a, Vasileios C. Pezoulas^b, Dimitrios I. Fotiadis^{b,c}, Vassilis Aidinis^{a,*}

^a Institute for Fundamental Biomedical Research, BSRC Alexander Fleming, Vari GR16672, Greece

^b Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, Ioannina GR45110, Greece

^c Biomedical Research Institute, FORTH, Ioannina GR45110, Greece

ARTICLE INFO

Article history:

Received 23 December 2022

Received in revised form 23 March 2023

Accepted 24 March 2023

Available online 25 March 2023

Keywords:

Idiopathic pulmonary fibrosis (IPF)

Machine learning

Diagnostic biomarkers

Omics data

ABSTRACT

Pulmonary fibrosing diseases are in the very epicenter of biomedical research both due to their increasing prevalence and their association with SARS-CoV-2 infections. Research of idiopathic pulmonary fibrosis, the most lethal among the interstitial lung diseases, is in need for new biomarkers and potential disease targets, a goal that could be accelerated using machine learning techniques. In this study, we have used Shapley values to explain the decisions made by an ensemble learning model trained to classify samples to an either pulmonary fibrosis or steady state based on the expression values of deregulated genes. This process resulted in a full and a laconic set of features capable of separating phenotypes to an at least equal degree as previously published marker sets. Indicatively, a maximum increase of 6% in specificity and 5% in Mathew's correlation coefficient was achieved. Evaluation with an additional independent dataset showed our feature set having a greater generalization potential than the rest. Ultimately, the proposed gene lists are expected not only to serve as new sets of diagnostic marker elements, but also as a target pool for future research initiatives.

© 2023 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive scarring disease of the lungs affecting mainly elder males. Patients present progressive decline of respiratory processes leading to dyspnea, fatigue and ultimately respiratory failure [1]. Epidemiologically it is the most deadly of the interstitial lung diseases (ILD) [2] affecting a great number of individuals in an annual base [3]. Although numerous research deliverables have focused on IPF, both its etiology and more importantly a curative option remain elusive. Currently, as SARS-CoV-2 infection has been suggested to stimulate the expression of pro-fibrotic targets [4] and interstitial lung disease patients present an increased risk of poor COVID-19 outcome [5,6], the proposal of a robust set of disease biomarkers and potential new targets has become more crucial.

Moving towards that direction, our laboratory recently created Fibromine, a collection of manually curated and consistently

processed IPF-related omics datasets [7]. Through Fibromine, the user can explore, mine and combine transcriptomic and/or proteomic datasets of both human and mouse origin, in search for interesting latent targets or validation of wet laboratory findings. Although combination of deregulated features across datasets in Fibromine is designed to select the most promising elements between any selected set of phenotypes, consensus features reported are sometimes numerous, rendering the selection of a small number of targets rather subjective and time consuming.

Machine learning techniques have been used many times in the past to analyze huge volumes of biomedical data in search for imperceptible, yet important patterns. Cancer research has benefited greatly from artificial intelligence-based drug or biomarker discovery studies [8], while such techniques have also led to the proposal of new biomarkers in the fibrotic context of non-alcoholic fatty-liver disease [9]. In addition, thanks to these *in silico* methodologies, disease endotypes requiring distinct treatment protocols have been revealed in numerous contexts, including that of acute respiratory syndrome [10]. As far as IPF research is concerned, most machine learning applications aim on stratifying patients based on

* Corresponding author.

E-mail address: aidinis@fleming.gr (V. Aidinis).

imaging data, such as in [11], while others exploit gene expression for recognition of distinct pathological states [12,13].

In this publication, we have used a semantics similarity/Shapley values-explained machine learning pipeline so as to prioritize and propose a new set of IPF diagnostic biomarkers. More specifically, ensemble learning models were successfully trained to discern pathological from steady-state samples according to expression data fetched from Fibromine. Shapley (Shapley Additive exPlanations; SHAP) values were then used to explain model decisions and rank/select 76 features as the most diagnostically valuable. Text mining and functional characterization of the selected genes revealed both well-established and slightly researched in the IPF context features, which constitute a valuable source of biomarkers and potential disease targets, respectively. Different ranking aggregation methods were used to integrate results across multiple models and select a shorter, lite version of the 76 Shapley-prioritized features. Comparison of machine learning-guided classification of samples based on either our lite or other already published biomarker sets revealed a wider applicability potential of the former. Last, interesting correlations were revealed between certain prioritized features and available spirometry measurements, further supporting the importance of the markers reported.

2. Materials and methods

2.1. Semantics similarity prioritization

Fibromine online interface was used to fetch consensus IPF_vs_Ctrl differentially expressed genes across seven datasets (GSE10667, GSE24206, GSE48149, GSE47460_GPL6480, GSE53845, GSE83717, GSE99621) selected based on a set of benchmarking characteristics [7]. Default thresholds for the identification of consensus deregulated genes were applied (absolute fold change of at least 1.2 on natural scale and p-value less than 0.05). Significant differential expression towards the same direction (up or down) in at least four out of the aforementioned seven datasets was required for a gene to be considered consensus deregulated. The repeated and independent nature of these findings greatly reduces the chances of false positive hits, especially for top deregulated features, similarly to past smaller scale attempts [14]. As a single value representation of the differential gene expression patterns reported across datasets, consensus fold change ($FC_{consensus}$) was used [7]. For GSE24206, Advanced IPF samples were used as the pathological ones. Subsequently, in order to reduce the chances of model overfitting (2.3. Machine learning models tuning, training and evaluation) while maintaining vastly trustworthy elements, the genes most related to the term *pulmonary fibrosis* (DOID:3770) were selected via semantics similarity prioritization performed by the R Bioconductor package DOSE (v.3.16.0) [15]. The per feature yielded similarity score (SS) was multiplied by $FC_{consensus}$ to obtain an integrated score $S_i = SS * FC_{consensus}$, sharing information between differential expression and fibrosis-association. S_i was used to rank genes in fashion similar to [16]. Top 100 up and top 100 down-regulated genes were isolated for downstream analysis.

2.2. Preprocessing of Fibromine gene expression data

Normalized gene expression values of the above datasets retrieved through Fibromine *Download* tab were used as input to machine learning models. More specifically, per dataset normalized values were standardized using vanilla R functions so as to obtain normalized expression values within a unified numerical range. Missing values were replaced by zero. The same process was applied to a couple of big datasets (GSE32537, GSE47460_GSE14550) not implicated in the training-testing procedure, but used as an external validation sample set (2.4. Machine learning models tuning, training

and evaluation). Common genes across all aforementioned datasets were isolated and then intersected with the 200 ones prioritized by semantics similarity (2.1. Semantics similarity prioritization). A final list of 172 features was used as the training set of machine learning models (Table S1).

2.3. Assessment of sex-specific feature expression

To assess any sex-specific effect on the expression of the selected 172 features, samples hierarchical clustering using Euclidean distance and complete linkage was performed. Sex metadata were retrieved from respective GEO entries and/or publications whenever available (62.5% of the training/testing and 100% of the external evaluation cases). For clustering purposes, preprocessed expression data were used, while samples without any sex annotation were removed prior to clustering.

2.4. Machine learning models tuning, training and evaluation

XGBoost models with a *binary:hinge* objective were tuned, trained and tested using the python libraries sklearn (v.1.0) [17] and xgboost (v.1.5.2) [18]. XGBoost (Extreme Gradient Boosting) is a scalable implementation of the gradient tree boosting supervised machine learning technique [18]. The latter, an ensemble learning method, is proven to produce high-quality results not only under standard [19] but also realistic circumstances [20].

The model was trained/tested using a Monte Carlo cross validation (MCCV) approach with a 75:25 train:test split iterated ten times to account for differences during data splits. To avoid overfitting, apart from reducing the number of features (2.1. Semantics similarity prioritization), per model hyperparameter tuning was performed using a stratified 5-fold approach with random shuffling on each training set followed by a grid-search cross-validation strategy. The hyperparameters tuned were: $n_estimators$ (the number of decision trees to use), $learning_rate$ (also known as *eta*; a regularization parameter adjusting feature weights during boosting in a step-wise manner), max_depth (the maximum depth of each tree), reg_alpha (the L1 regularization parameter that corresponds to Manhattan distance) and reg_lambda (the L2 regularization parameter that corresponds to squared Euclidean distance).

Accuracy, precision, sensitivity and specificity, were used for the evaluation of the models' performance. In addition, F1-score was applied to assess the precision-sensitivity tradeoff of each model and Matthews correlation coefficient (MCC) was calculated as below to summarize the per model confusion matrix in a single value. The median across models value for each evaluation metric was reported.

$$MCC = \frac{(TP * TN - FP * FN)}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}}$$

Models were also evaluated using GSE32537 and GSE47460_GPL14550, two datasets not implicated in model training-testing. The whole machine learning pipeline is visualized in Fig. S1.

2.5. Comparison of biomarker lists

XGBoost was employed to compare the capability of different biomarker lists for phenotype separation. The same samples used for shaping the here proposed lists were exploited for training of all models, while the intersection of the already-published marker sets to those genes shared by the Fibromine-retrieved datasets were used as features (Fig. S4). Models were tuned, trained, and tested using the previously described pipeline (Fig. S1).

2.6. Machine learning model explanation

Game theory inspired Shapley values were calculated using the homonym python library (v.0.39.0) (<https://github.com/slundberg/shap>). Per model mean absolute SHAP values were calculated, as proposed previously [21,22], to assess the importance of individual features in diagnosing the IPF phenotype. Features with a zero absolute SHAP value were filtered out. SHAP values are able to explain machine learning models both locally and globally in a consistent and accurate fashion [23]. Briefly, SHAP values reflect the contribution of each feature on the final prediction of the model after taking into consideration all possible feature combinations (coalitions) formed one feature at a time and then averaging the marginal contributions of each feature in a weighted manner. Collectively, given an input subset of features $\{d_1, d_2, d_3, \dots, d_M\}$ from a larger set of L features $\{d_1, d_2, d_3, \dots, d_L\}$ where $M \leq L$, the SHAP value of a feature $d_j \in D (S_j)$ is calculated as thus:

$$S_j = \sum \frac{|D|!(P - |D| - 1)!}{P!} (f_d(D \cup \{d\}) - (f_d(D)))$$

where $|D|$ is the number of features in D and $f_d(D)$ is the expected value of the function conditioned on P .

2.7. Pathway analysis

R Bioconductor package clusterProfiler (v.3.18.1) [24] was used for over-representation analysis of the important (non-zero absolute SHAP) machine learning features. Gene ontology terms of all three categories (*Cellular Component*; *Molecular Function*; *Biological Process*) were auto-retrieved from the package, while MSigDB hallmark gene sets were fetched from the R package msigdb (v.7.5.1) (<https://igordot.github.io/msigdb/>).

2.8. Text mining

PubMed 2022 baseline was fetched from the respective ftp site and XML R package (v.3.99.0.8) was used to form an abstract corpus. The latter was queried for the elements containing *IPF[All Fields] OR ("pulmonary fibrosis"[MeSH Terms] OR "pulmonary fibrosis"[All Fields]) OR ("lung diseases, interstitial"[MeSH Terms] OR "interstitial lung diseases"[All Fields] OR "interstitial lung disease"[All Fields])* using rentrez R package (v.1.2.3). Human gene atomization based on HGNC symbols was performed with pubmed.mineR package (v.1.0.19) [25] and the results were intersected with the machine-learning - SHAP value prioritized gene list.

2.9. Ranking aggregation methods

SHAP-weighted majority voting, an ad hoc modified version of majority voting in which genes are ranked by decreasing gene-wise Shapley value weights was applied. More specifically, the importance, and thus final rank, of any given gene i was calculated as the sum of the per model (m) ratio of the total number of non-zero mean absolute SHAP value features in the model (n_m) to the rank of the gene in the model (r_m) according to decreasing order of SHAP values:

$$\sum_{m=1}^M \frac{n_m}{r_m}$$

Two well established best performers were also applied. MAIC ranking aggregation algorithm was retrieved from the respective github repository (<https://github.com/baillielab/maic>), while BIRRA script was fetched from (https://github.com/baillielab/comparison_of_RA_methods). During MAIC and BIRRA ranking aggregation, each model was treated as a different category. The results of the above

three methods were compared by means of Kendall correlation as implemented in the homonym R package (v.2.2.1) and function.

2.10. Correlation analysis

Normalized gene expression values of selected machine learning prioritized features were retrieved from Fibromine and correlated with spirometry measurements retrieved using GEOquery package (v. 2.58.0) [26]. Both expression and spirometry values originated from datasets GSE47460_GPL6480, GSE47460_GPL14550 and GSE32537. Correlation included those dataset samples used during machine learning training/testing not having a missing value for each spirometry measurement examined. Spearman correlation was separately performed for each dataset using *cor.test* vanilla R (v.4.0.4) function. Relationships with an absolute rho value of at least 0.5 and a p-value less than 0.05 were considered as statistically significant strong.

3. Results

3.1. XGBoost machine learning algorithm successfully separates IPF from control individuals

The most important differential expression comparison hosted in Fibromine, both due to its anthropocentric character and to the number of available datasets, is that of IPF versus control individuals (*IPF_vs_Ctrl*). Among those datasets, seven were accredited with at least four stars by our datasets benchmarking strategy [7] and were thus distinguished as the most "reliable" combination set (*2.1 Semantics similarity prioritization*). Collectively, these datasets gather samples from 184 individuals, while their combination using Fibromine's *Datasets explorer* with default thresholds (*2.1 Semantics similarity prioritization*) reports 2182 consensus deregulated genes. Manually selecting biomarkers and/or targets from this big, yet biologically reasonable number of genes is a rather impossible task. Instead, as machine learning algorithms can provide an alternative solution to the aforementioned problem [27], we decided to create an explainable diagnostic model and assess the most valuable of the machine-used features.

Initially, to decrease the chances of a downstream model overfitting event due to the larger ratio of features (genes) to examples (lung samples) ($p \gg n$), we applied a semantics similarity prioritization technique described by [16] after some context-required modifications (*2.1 Semantics similarity prioritization*). Briefly, an ad hoc calculated score (S_i) sharing information between semantics similarity and gene expression was used to rank the 2182 genes and the top 100 up- and 100 down-regulated ones were selected for model training (Fig. 1A).

Next, normalized gene expression data were retrieved from Fibromine. To be suitable for any machine learning application, these data underwent several preprocessing steps (Fig. 1A). Initially, as selected datasets originated from different platforms, gene expression values were standardized to fit within the same numerical range. Subsequently, common across datasets features (see 2.2. *Preprocessing of Fibromine gene expression data*) were selected and then intersected with the previously selected top 200 semantics similarity prioritized genes. Eventually, our machine training dataset consisted of 184 examples (124 patients and 60 controls) and 172 features (Table S1) equally distributed between over and under expressed gene groups (84 up and 88 down regulated genes). Because IPF is known to affect males more often than females [28] and gene expression can be regulated in a sex-specific manner [29] hierarchical clustering was used to assess any differences between sexes for the 172 selected genes. No case of sex-specific expression was observed (Fig. S2).

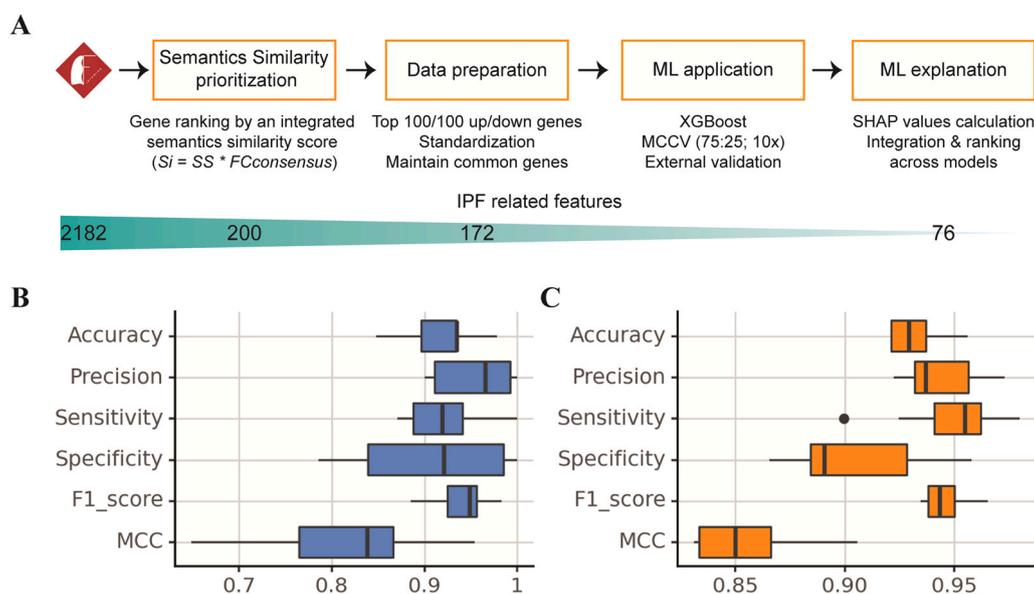


Fig. 1. An explainable machine learning based pipeline was successfully used for phenotype classification based on gene expression values. A. Pipeline's graphical abstract. Fibromine-derived, semantics similarity prioritized *IPF_vs_Ctrl* consensus differentially expressed genes were used for XGBoost modeling. The latter was explained by Shapley (SHAP) values. B. Model evaluation using the testing dataset. The trained models demonstrate very good performance. C. Model evaluation using an independent evaluation dataset. The models present the same if not better performance in a totally independent dataset. MCCV: Monte Carlo Cross Validation.

Table 1

Comparison of performance metrics between our full and lite model with those based on previous gene lists. The median across ten tuning-training-testing iterations is presented. Two significant digits have been maintained for all values. Top values are marked in blue and orange according to the evaluation data set. Full: the full model of this study; Lite: the lite model of this study; Test: testing set of the split; Val: external validation set.

Model	Performance metrics											
	Accuracy		Precision		Sensitivity		Specificity		F1-score		MCC	
	Test	Val.	Test	Val.	Test	Val.	Test	Val.	Test	Val.	Test	Val.
Full	0.93	0.93	0.97	0.94	0.92	0.95	0.92	0.89	0.95	0.94	0.84	0.85
Lite	0.93	0.94	0.97	0.95	0.94	0.95	0.91	0.92	0.95	0.96	0.85	0.88
Wan	0.93	0.93	0.95	0.95	0.94	0.97	0.89	0.87	0.95	0.95	0.84	0.86
Li	0.91	0.93	0.95	0.93	0.91	0.96	0.90	0.89	0.94	0.95	0.80	0.86
Zhu	0.92	0.92	0.95	0.92	0.94	0.94	0.89	0.87	0.95	0.93	0.81	0.83
Cui	0.93	0.92	0.97	0.92	0.94	0.96	0.91	0.85	0.96	0.94	0.85	0.83
Xu	0.91	0.86	0.94	0.85	0.95	0.93	0.86	0.74	0.94	0.89	0.80	0.69

XGBoost ensemble algorithm was used to classify IPF and control individuals [18]. Selection of this fast gradient boosting algorithm was made due to its regularized learning objective function which renders it robust to overfitting and enables feature selection, as well as due to its previous successful use with gene expression data [30]. In addition, the ability of XGBoost not to be affected by multi-collinearity phenomena [31] further supported its selection, as top deregulated genes are by default expected to co-fluctuate. Model tuning/training were performed with a ten-iteration MCCV strategy (2.4. Machine learning models tuning, training and evaluation) (Fig. 1A). To summarize model testing, across iterations the median value for each evaluation metric is reported. Specifically, XGBoost showed a very good performance in phenotype prediction as

summarized across the 10 models, with all metrics except MCC having a median of at least 0.9 (Fig. 1B; Table 1).

The efficiency of the trained models was further evaluated on two additional Fibromine-retrieved datasets (GSE32537, GSE47460_GPL14550) (Figs. 1A; S1). This data collection was processed simultaneously with the one used during tuning-training-testing, in order to obtain the exact same features (2.4. Machine learning models tuning, training and evaluation). The preprocessing pipeline yielded a dataset of 199 IPF and 119 control cases. As can be seen from Fig. 1C and Table 1, all models present a very good performance in this external dataset, very similar to that of the initial testing sets in addition to smaller fluctuations of the metric values of the former compared to the latter.

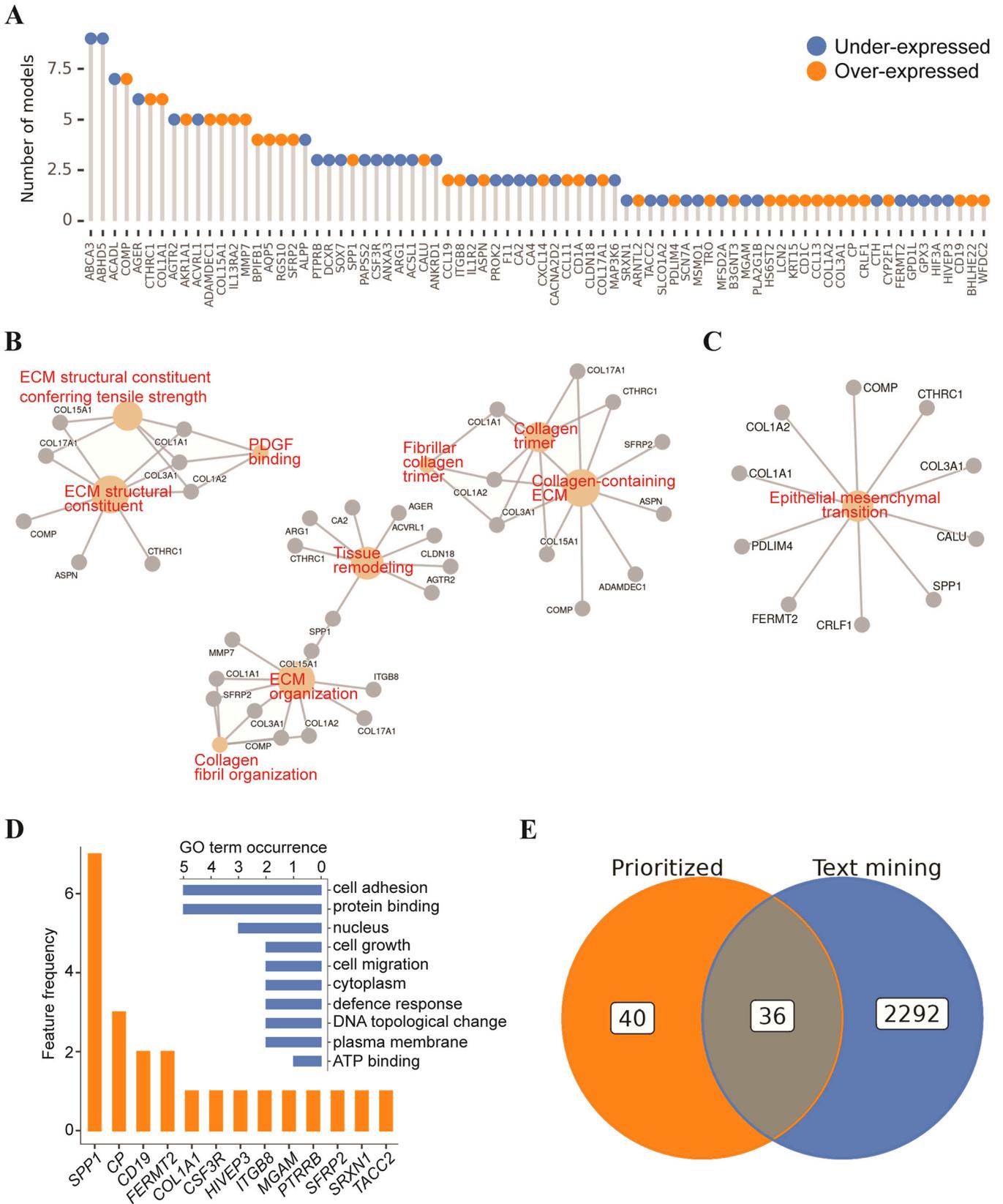


Fig. 2. Functional characterization of the prioritized list of features reveals their fibrosis-related nature. A. Features exploited by the XGBoost classifiers for the separation of IPF and control individuals. Direction of deregulation as established by $FC_{\text{consensus}}$. B. Pathway analysis based on Gene Ontology database. Molecular function, biological process and cellular component top enriched elements are visualized in different networks. C. Pathway analysis based on MSigDB. D. Prioritized features found in Corum3.0 protein-protein complexes and top GO terms characterizing them. E. Venn diagram of the intersection between text mining reported and prioritized list of genes. Almost half of the prioritized features have been highlighted in the past in relation to pulmonary fibrosis and interstitial lung diseases.

To conclude with, we were able to separate IPF from control individuals by applying a supervised machine learning algorithm on Fibromine gene expression data previously selected via semantics similarity.

3.2. SHAP values-based selection and characterization of ML used genes

In order to explain the decisions made by the models trained, we then calculated per model Shapley values for each one of the features. Non-zero mean absolute SHAP values revealed a varying number of per model important genes with a minimum of 11 and a maximum of 32 significant features (Fig. S3). Union of the features across the 10 models led to a total of 76 non-zero SHAP genes, with several of them characterized by a repeated and others by a less frequent pattern of emergence (Fig. 2A). For the sake of brevity, this marker set will be called *long* or *full* list for the rest of the paper.

To functionally characterize this full marker list, we initially performed pathway analysis using Gene Ontology (GO) database. Top enriched terms revealed fibrosis related entities such as the biological processes of *collagen fibril organization*, *extracellular matrix organization*, molecular functions including *tissue remodeling* and cellular components such as *fibrillar collagen trimer*, *collagen containing ECM* (Fig. 2B; Table S2). The same analysis was repeated using the MSigDB hallmark gene sets as a reference database, revealing that our gene list is enriched for *epithelial-mesenchymal transition* features (Fig. 2C; Table S3), known to be implicated in pulmonary fibrosis [32]. Moreover, CORUM3.0 [33] query using the full feature list implicated many of them with protein complexes related to the plasma membrane, cell adhesion and migration, immune response and integrin-mediated signaling pathways (Fig. 2D; Table S4). Last, text mining of PubMed abstracts selected using the keywords *pulmonary fibrosis*, *IPF* or *interstitial lung diseases* revealed a set of 36 genes already mentioned in the context of lung fibrosis (Fig. 2E; Table S5).

A closer inspection of the SHAP-selected features revealed that several of them are already known to be implicated in lung fibrosis. For instance, *AGER*, *SPPI1*, *MMP7* and *CRLF1* are included in a review list of transcriptionally deregulated genes during IPF compiled by [34]. In addition, genes such as *COL1A1*, *COL15A1*, *COMP*, *MMP7* and *LCN2* were identified as potential IPF drivers by previous exploratory research deliverables comparing the transcriptomes of IPF patients to those of control individuals (Table 2) [35–39].

Collectively, a great number of the 76 genes actually used during machine learning phenotype classification are known to be implicated in pulmonary fibrosis and 12 of them were even proposed as IPF diagnostic biomarkers at least once in the past. These observations support the fibrosis-related character of our list. More interestingly, half of the machine learning-chosen genomic elements have not been extensively studied in the context of lung fibrosis as proposed by PubMed abstracts mining, thus consisting a pool of potentially significant targets for downstream experimentations.

3.3. SHAP-based ranking aggregation of prioritized features

To obtain a unique robust set of important features sharing information across all ten XGBoost models, we integrated SHAP-ordered lists using three ranking aggregation methods (Fig. 3A). BIRRA [40] and MAIC [41], two of the best strategies in handling large numbers of ranking sources with high or unknown quality of heterogeneity [42] were applied along with *SHAP-weighted majority voting*, an ad hoc crafted method extending the voting ranking aggregation method (2.9. *Ranking aggregation methods*). Interestingly, although each method proposed a different ranking of the features in the full biomarker list (Fig. 3B), all three strategies seem to have comparable performance as they share 65% of the 1st quartile of the features list. More than half of these common features (8/13) are

among the text mining highlighted ones, while four (*COMP*, *COL1A1*, *COL17A1* and *COL5A2*) have been proposed as biomarkers in the past [35,36,38,39]. In addition, comparison of the lists along their whole length using Kendall ranking correlation, revealed strong similarities of BIRRA with each of the SHAP-weighted majority voting and MAIC methods, while the latter two share less common rankings (Fig. 3C).

The aforementioned data support that ranking aggregation of Shapley value-selected features across model iterations can lead to an integrated list of informed ordering. Having various levels of representation in the bibliography, top genes of the aggregation methods comprise an interesting collection for further examination.

3.4. Comparison with previous biomarker lists

Subsequently, we wanted to compare our full feature list with already published, *in silico* predicted lists of markers separating IPF from control cases [35–39]. Regardless of the much larger size of our set, a complete intersection with any of the previous lists was not achieved (Fig. 4A; Table 2). Moreover, the already available marker lists were greatly heterogeneous (Fig. 4A).

Motivated by the different nature of the aforementioned gene collections we examined their ability of distinguishing phenotypes relatively to that of our full feature list (Fig. 2A), using XGBoost as a common baseline methodology (2.5. *Comparison of biomarker lists*). Furthermore, in order to propose a laconic set of IPF biomarkers along with the full feature list, we trained and evaluated another 10 models using the 13 of the top 20 features shared between the three ranking aggregation methods applied to the full biomarker list (Fig. 3B).

As can be seen from Fig. 4B and Table 1, when models are evaluated on testing data, our lite model is the best performer along with the models based on Wan and Cui biomarkers. Their most of the times equal performance is improved by 2–3% and even 5% compared to the rest of the models. Xu and colleagues based algorithm has the least favorable performance, with the exception of the sensitivity metric. On the other hand, when external evaluation datasets are examined (Fig. 4C; Table 1) our lite model has the best performance in all but the sensitivity measurement, an observation that suggests a better generalization potential of the herein shaped laconic marker set. More specifically, the lite model achieved a 92% specificity followed by that of Li and our full model, which scored three units less. Similarly, our lite model was characterized by the best accuracy (94%), MCC (88%) and F1-score (96%), along with the best precision (95%) shared with the Wan model. Xu-based phenotype classifier had the least favorable performance on the external validation dataset.

SHAP value prioritization of the lite model features followed by ranking aggregation using the ad hoc crafted method was applied (Fig. 4D). As expected, genes already known to be fibrosis-implicated such as *MMP7*, *COL15A1* and *COMP* [34] (Table 2) have been ranked on the top of the list by our ad hoc designed aggregation method. *IL13RA2*, a gene neither reported as an IPF biomarker in the past nor found by text mining was ranked fourth. Correlation analysis of the lite model's features normalized expression with spirometry measurements available in three datasets used during machine learning led to the validation of known and the establishment of latent relationships (Table S6). Indicatively, *IL13RA2* was strongly negatively correlated to the % predicted forced vital capacity (FVC) in two out of three cohorts (Fig. 4E), while the opposite pattern was observed for *PAPSS2* (Fig. 4F).

4. Discussion

IPF is a progressive lethal disease that is expected to affect a growing percentage of individuals, as societies of especially the developed world are getting elder [43,44]. Unfortunately, despite

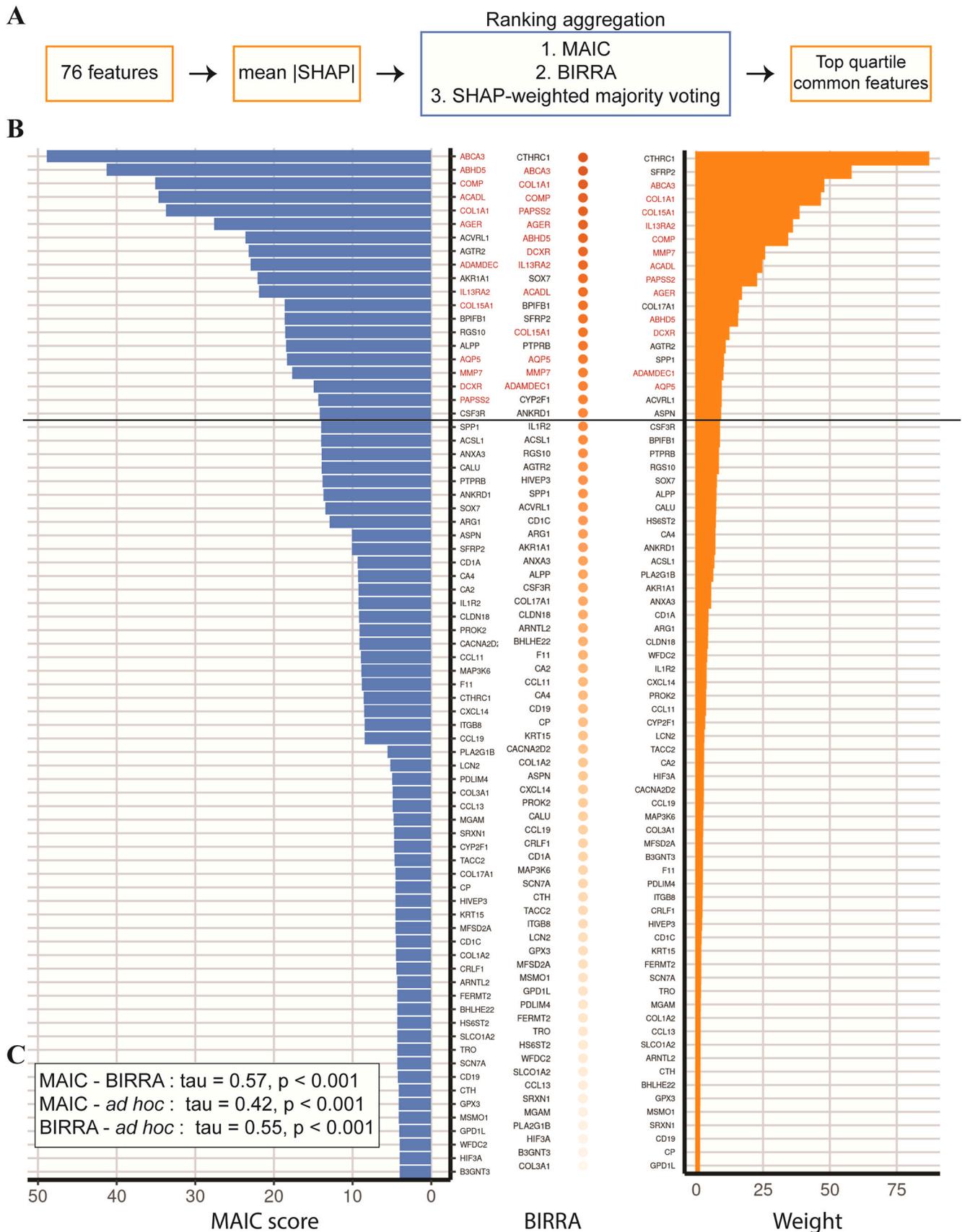


Fig. 3. Ranking aggregation methods were used to integrate gene rankings from each model iteration. A. Basic pipeline. Two well established (BIRRA, MAIC) and an ad hoc developed method (SHAP-weighted majority voting) were applied. B. Prioritized features ranking post to aggregation of the individual models. Each method returns a distinct ranking. Genes marked with red are those in the top 20 (~1st quartile) shared between the rankings. C. Kendall correlation of the rankings. BIRRA results are strong correlated with those of the other two methods.

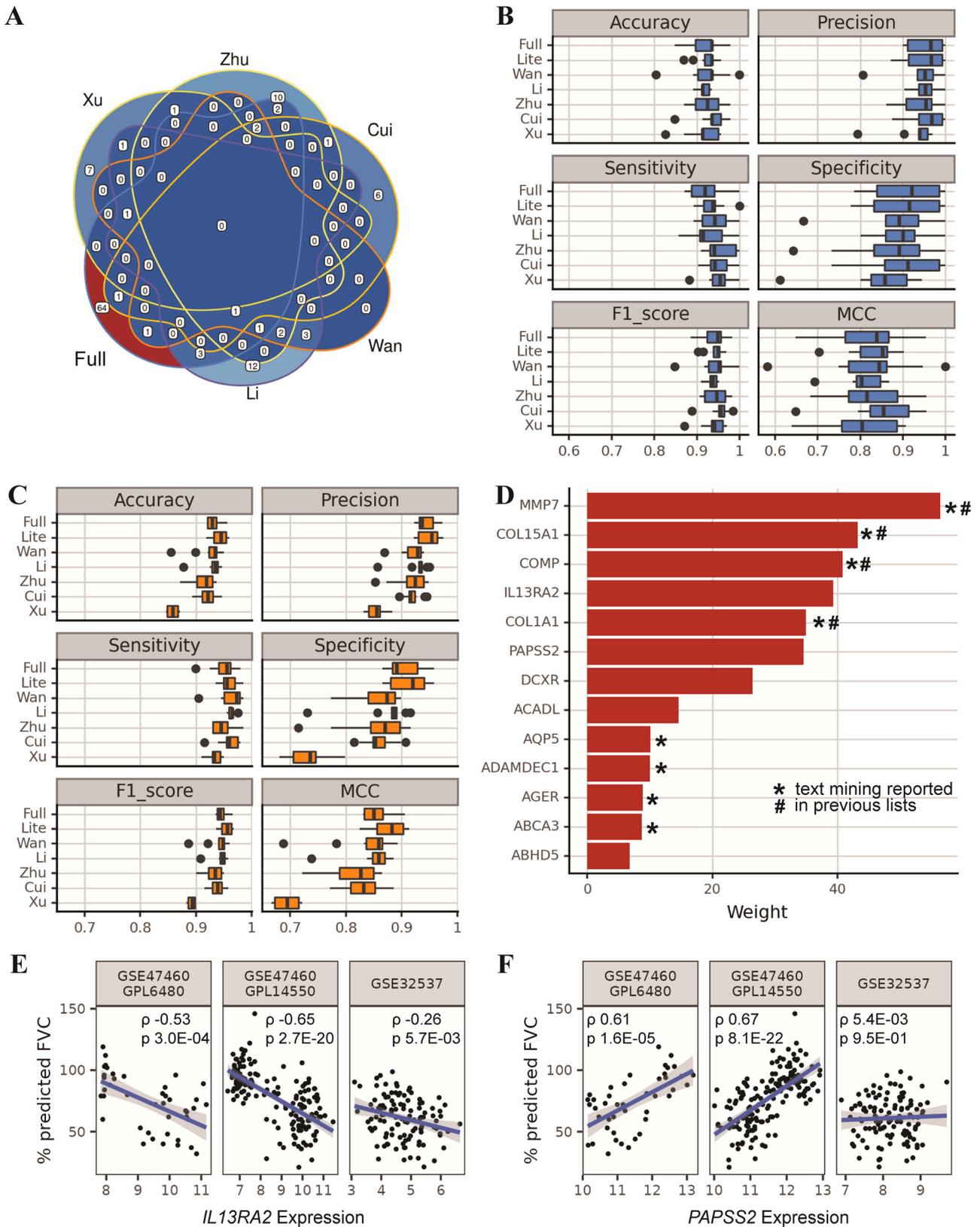


Fig. 4. Comparison of a lite model with those of previous biomarker lists shows greater generalization potential. A. Intersection of the prioritized biomarkers with those of previous publications. Great heterogeneity is observed among lists. B. Model evaluation using the testing dataset. The herein introduced model is among the top performers. C. Model evaluation using an independent evaluation dataset. The proposed lite model shows a greater potential of generalization compared to the other methods. D. Integration of the ten lite models using the ad hoc crafted method. The list comprises of both known and less explored pulmonary fibrosis targets. E-F. Correlation of *IL13RA2* and *PAPSS2* expression values with those of predicted Forced Vital Capacity (FVC) in three datasets. Strong significant relationship is established in two out of three cohorts for both genes.

Table 2

Prioritized features found in previous idiopathic pulmonary fibrosis versus control differential expression-based biomarker lists.

Prioritized Gene	Xu et al. PMID: 32899090	Zhu et al. PMID: 34522168	Cui et al. PMID: 33672678	Wan et al. PMID: 34476240	Li et al. PMID: 32782692
COL1A1		X	X	X	X
COL15A1					X
COMP		X			
MMP7		X	X	X	
COL17A1					X
SPP1		X	X	X	
ASPN				X	
LCN2		X			
COL3A1		X		X	X
ITGB8			X		
COL1A2	X			X	X
CP					X

decades of intensive research, the factors responsible for IPF manifestation and establishment remain largely unknown. In addition to the latter, COVID-19 pandemic and pulmonary fibrosis seem to be ominously interconnected [4,6] rendering the discovery of new targets extremely urgent. Here, we have exploited Fibromine, a database and mining tool of manually curated, consistently analyzed pulmonary fibrosis omics datasets in an ensemble learning-powered attempt to *in silico* highlight previously neglected IPF-implicated molecules.

The rationale behind this study was to train and use supervised machine learning models for samples binary classification on an either pathological or steady state, and to subsequently evaluate the data-driven decisions of the models as they are expected to reflect important homeostatic changes. The ultimate goal was to obtain a new set of disease targets capable to inform future wet laboratory experimentations. For this purpose, the most similar *IPF_vs_Ctrl* lung Fibromine datasets were chosen and their consensus top deregulated genes were submitted to semantics similarity analysis. The repeated and independent nature of the selected findings combined with their utmost ranking among the deregulated elements greatly limits the chances of a false positive hit, similarly to [14]. Ultimately, via this semantics similarity feature selection we have reduced the probability of model overfitting by reducing the number of available features, while at the same time we maintained an objective set of important fibrosis-related terms.

Based on the selected gene expression data, an XGBoost ensemble learner was successfully tuned, trained and tested for separating patients from pulmonary fibrosis free individuals using an MCCV approach. The XGBoost model was specifically chosen for several reasons including its previous effective use with gene expression data [30], as well as its resistant nature to overfitting and multi-co-linearity phenomena [31]. The latter algorithmic property is extremely significant as it enables the maintenance and downstream prioritization of equally significant deregulated features which may belong to distinct molecular processes; features that would have been removed prior to any other multi-co-linearity susceptible approach. Models evaluation using the testing set returned a median of at least 0.92 for all but one metrics, while model assessment with an entirely independent couple of Fibromine datasets resulted in an at least equally good classification, scoring a median of minimum 0.85 and maximum 0.95 in the same performance metrics.

Subsequently, per gene SHAP values were calculated in order to explain features contribution to each of the models. 76 genes with at least one non-zero SHAP value in any of the 10 iterations were selected for downstream processes. Functional characterization based on pathway analysis and CORUM3.0 mining revealed several fibrosis-related entities and protein complexes supporting the fibrotic-related character of the selected genes, respectively. In

addition, almost half of them were revealed to be associated with pulmonary fibrosis and/or interstitial lung diseases based on PubMed text mining. More importantly, 12 of those machine learning-selected features were suggested by previous publications as potential IPF biomarkers, further strengthening the fibrotic character of our list.

To aggregate in a single list the 76 features collectively proposed by the ten models, two of the best performing [42] - BIRRA and MAIC - and an ad hoc developed ranking aggregation methods were applied. Based on the popular majority voting strategy, the herein formulated ranking aggregation method delivered very similar rankings with the other two as revealed by Kendall correlation. In addition, 13 of the 20 top ranking genes were common across all lists and more than half of those genomic elements were underlined as fibrosis-associated via text mining. Importantly, four of them (*COMP*, *COL1A1*, *COL17A1* and *COL5A2*) were already included in previous lists of *in silico* predicted IPF biomarkers [35,36,38,39].

Last, an attempt was made to compare our full set of proposed markers with *in silico*-formed ones from previous publications that exploited publicly available datasets in a mostly network-oriented manner [35–39]. For such a purpose, ten XGBoost models were tuned and trained for each of the already published biomarker lists following the exact same pipeline as in our full model. In parallel, an extra model was trained based the top 13 genes shared by all ranking aggregation methods applied on the full feature list (Fig. 3B), with the aim of presenting a smaller set of similar size to that of the already available marker lists.

As expected based on past machine learning applications on gene expression data [45], juxtaposition of the models' performance showed relatively small differences. In brief, evaluation on the testing set placed our lite model among the top performing methods, while Xu-based classifier showed the lowest efficiency in separating IPF from control individuals. The latter observation can be potentially attributed to the formation of the initial Xu list using only a couple of datasets [37] possibly not adequate to properly address IPF heterogeneity. Model performance on the external set of data suggested that our lite feature list has a greater generalization potential, with only its sensitivity being outperformed by just a couple of models. Thus, the above results propose the aforementioned 13 genes as a robust set of IPF biomarkers.

In addition to introducing a new set of diagnostic genes able to classify IPF and healthy cases, this endeavor also attempts to bring into scientific attention several IPF-related features overlooked by previous *in silico* studies. For example, focusing on the genes of our lite model, apart from well-known pulmonary fibrosis players, such as *MMP7* [46–48], *COL1A1* [47,49,50], *COL15A1* [47], *COMP* [51] and *AGER* [52,53], there are features that have not been extensively studied in pulmonary fibrosis. Although, interleukin 13 receptor subunit alpha 2 (*IL13RA2*) has been evidenced to mediate pro-fibrotic *TGFb-1* signaling in a colitis model [54] and was here correlated with several pulmonary function tests, its potential implication in pulmonary fibrosis needs to be thoroughly investigated. Other genes unexplored in the lung scarring context include *PAPSS2* which has been recently implicated in pro-tumorigenic processes of cutaneous squamous cell carcinoma cells [55], *DCXR* whose suppression has been shown to ameliorate renal fibrosis [56] and *ACADL* which has been evidenced to offer an at least indirect protection against steatosis [57]. All above three genes were here shown to have significant correlations with lung function physiological measurements.

Furthermore, other interesting genes lie among the most important features of our full model. Secreted frizzled-related protein 2 (*SFRP2*) which is among the top 20 prioritized genomic elements of both BIRRA and SHAP-weighted majority voting is a known marker of myofibroblast progenitors in systemic skin sclerosis [58] and has also been recently implicated in breast cancer, participating in a

lung-breast cancer axis of communication [59]. Nevertheless and although its transcription has been located in a fibroblast population by [60], its role in IPF has not yet been clarified. Even though reported by text mining, *CTHRC1* and *ABCA3* are two other features ranked among the top three genes of both BIRRA and the here developed ranking aggregation method, whose exact role in pulmonary fibrosis may worth a detailed dissection. More specifically, although known to be over-expressed during pulmonary fibrosis, *CTHRC1* has been only recently shown to mark cell types with pathological expression profile in the fibrotic lung [50]. More importantly, as far as we are able to know, reported *ABCA3* implication in pulmonary fibrosis has until now been restricted to mutations found in infant and adult IPF patients [61]. Nevertheless, the ILD phenotype of an infant was most recently reversed after disrupting an inherited *ABCA3* biallelic mutation via lung transplantation [62].

Our study is not without limitations. First, models constructed on previously published biomarker lists, included a big yet not complete set of per list proposed genes, as a comprehensive inclusion would have automatically differentiated their baseline training data from that of our full list. Second, PubMed baseline collection of 2022 publications although fairly up-to-date does not include the latest research deliverables and may have thus slightly affected our text mining results. Manual literature search results performed in an attempt to fill that publication gap, although thorough, are not expected to be impeccable.

Overall, this study proposed a full and a laconic set of marker genes able to separate IPF from pulmonary-fibrosis free individuals using a transparent machine learning pipeline. Starting from a long list of differentially expressed genes it offered a full and a lite set of IPF diagnostic features hosting both well-known and less investigated targets. Furthermore, the results of an ad hoc ranking aggregation method have been shown to reflect underlying biology in an adequate degree compared to two of the currently top-performing strategies. Future wet laboratory experiments are needed to establish the role of the here highlighted disease targets and assess the biomarkers effectiveness at the laboratory bench.

Funding

The research project was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “2nd Call for H.F.R.I. Research Projects to support Faculty Members & Researchers” (#3565 to VA). The funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

CRedit authorship contribution statement

Dionysios Fanidis: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Visualization. **Vasileios C. Pezoulas:** Conceptualization, Methodology, Writing – review & editing. **Dimitrios I. Fotiadis:** Writing – review & editing. **Vassilis Aidinis:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Prof. A. Tzioufas (Department of Pathophysiology, School of Medicine, University of Athens) for critical suggestions.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.03.043.

References

- [1] Raghu G, et al. Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 2018;198(5):e44–68.
- [2] Garcia CK. Idiopathic pulmonary fibrosis. *Proc Am Thorac Soc* 2011;8(2):158–62.
- [3] Olson AL, et al. The epidemiology of idiopathic pulmonary fibrosis and interstitial lung diseases at risk of a progressive-fibrosing phenotype. *Eur Respir Rev* 2018;27(150):180077.
- [4] McDonald LT. Healing after COVID-19: are survivors at risk for pulmonary fibrosis? *Am J Physiol Lung Cell Mol Physiol* 2020;320(2):L257–65.
- [5] Esposito AJ, et al. Increased odds of death for patients with interstitial lung disease and COVID-19: A Case–Control Study. *Am J Respir Crit Care Med* 2020;202(12):1710–3.
- [6] Drake TM, et al. Outcome of hospitalization for COVID-19 in patients with interstitial lung disease. An International Multicenter Study. *Am J Respir Crit Care Med* 2020;202(12):1656–65.
- [7] Fanidis D, Moulos P, Aidinis V. Fibromine is a multi-omics database and mining tool for target discovery in pulmonary fibrosis. *Sci Rep* 2021;11(1):21712.
- [8] Huang S, et al. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genom Proteom* 2018;15(1):41.
- [9] Han N, et al. Identification of biomarkers in nonalcoholic fatty liver disease: a machine learning method and experimental study. *Front Genet* 2022;13.
- [10] Calfee CS, et al. Acute respiratory distress syndrome subphenotypes and differential response to simvastatin: secondary analysis of a randomised controlled trial. *Lancet Respir Med* 2018;6(9):691–8.
- [11] Nishikiori H, et al. Validation of the artificial intelligence software to detect chronic fibrosing interstitial lung diseases in chest X-ray. *Eur Respir J* 2021;58(Suppl. 65):SOA1211.
- [12] Choi Y, et al. Analytical performance of Envisia: a genomic classifier for usual interstitial pneumonia. *BMC Pulm Med* 2017;17(1):141.
- [13] Kim SY, et al. Classification of usual interstitial pneumonia in patients with interstitial lung disease: assessment of a machine learning approach using high-dimensional transcriptional data. *Lancet Respir Med* 2015;3(6):473–82.
- [14] Melboucy-Belkhir S, et al. Forkhead Box F1 represses cell growth and inhibits COL1 and ARPC2 expression in lung fibroblasts in vitro. *Am J Physiol Lung Cell Mol Physiol* 2014;307(11):L838–47.
- [15] Yu G, et al. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* 2015;31(4):608–9.
- [16] Maghsoudloo M, et al. An efficient hybrid feature selection method to identify potential biomarkers in common chronic lung inflammatory diseases. *Genomics* 2020;112(5):3284–93.
- [17] Pedregosa F, et al. Scikit-learn: machine learning in python. *J Mach Learn Res* 2011;12:2825–30.
- [18] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. Association for Computing Machinery: San Francisco, California, USA; 2016. p. 785–94.
- [19] Li P. Robust LogitBoost and adaptive base class (ABC) LogitBoost. In: Proceedings of the twenty-sixth conference annual conference of uncertainty in artificial intelligence; 2010.
- [20] He X, et al. Practical lessons from predicting clicks on ads at facebook. *Adkdd'14* 2014:1–9.
- [21] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Proceedings of the 31st international conference on neural information processing systems. Curran Associates Inc.: Long Beach, California, USA; 2017. p. 4768–77.
- [22] Lundberg SM, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2020;2(1):56–67.
- [23] Lundberg SM, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* 2018;2(10):749–60.
- [24] Yu G, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS J Integr Biol* 2012;16(5):284–7.
- [25] Rani J, Shah AR, Ramachandran S. pubmed.mineR: an R package with text-mining algorithms to analyse PubMed abstracts. *J Biosci* 2015;40(4):671–82.
- [26] Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and bioconductor. *Bioinformatics* 2007;23(14):1846–7.
- [27] Li L, Ching W-K, Liu Z-P. Robust biomarker screening from gene expression data by stable machine learning-recursive feature elimination methods. *Comput Biol Chem* 2022;100:107747.

- [28] Raghu G, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011;183(6):788–824.
- [29] Gershoni M, Pietrokovski S. The landscape of sex-differential transcriptome and its consequent selection in human adults. *BMC Biol* 2017;15(1):7.
- [30] Dimopoulos AC, et al. Combining multiple RNA-Seq data analysis algorithms using machine learning improves differential isoform expression analysis. *Methods Protoc* 2021;4:4.
- [31] Parsa AB, et al. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accid Anal Prev* 2020;136:105405.
- [32] Rout-Pitt N, et al. Epithelial mesenchymal transition (EMT): a universal process in lung diseases with implications for cystic fibrosis pathophysiology. *Respir Res* 2018;19(1):136.
- [33] Giurgiu M, et al. CORUM: the comprehensive resource of mammalian protein complexes—2019. *Nucleic Acids Res* 2019;47(D1):D559–63.
- [34] Vukmirovic M, Kaminski N. Impact of transcriptomics on our understanding of pulmonary fibrosis. *Front Med* 2018;5:87.
- [35] Zhu K, et al. Integrated analysis of the molecular mechanisms in idiopathic pulmonary fibrosis. *Int J Med Sci* 2021;18(15):3412–24.
- [36] Cui Y, et al. Identification of key candidate genes involved in the progression of idiopathic pulmonary fibrosis. *Molecules* 2021;26:4.
- [37] Xu Z, et al. Using bioinformatics approach identifies key genes and pathways in idiopathic pulmonary fibrosis. *Medicine* 2020;99:36.
- [38] Li D, Liu Y, Wang B. Identification of transcriptomic markers for developing idiopathic pulmonary fibrosis: an integrative analysis of gene expression profiles. *Int J Clin Exp Pathol* 2020;13(7):1698–706.
- [39] Wan H, et al. Identification of hub genes and pathways associated with idiopathic pulmonary fibrosis via bioinformatics analysis. *Front Mol Biosci* 2021;8.
- [40] Badgeley MA, Sealfon SC, Chikina MD. Hybrid Bayesian-rank integration approach improves the predictive power of genomic dataset aggregation. *Bioinformatics* 2015;31(2):209–15.
- [41] Li B, et al. Genome-wide CRISPR screen identifies host dependency factors for influenza A virus infection. *Nat Commun* 2020;11(1):164.
- [42] Wang B, et al. Systematic comparison of ranking aggregation methods for gene lists in experimental results. *Bioinformatics* 2022;btac621.
- [43] Navaratnam V, et al. The rising incidence of idiopathic pulmonary fibrosis in the UK. *Thorax* 2011;66(6):462.
- [44] Hutchinson JP, et al. Increasing global mortality from idiopathic pulmonary fibrosis in the twenty-first century. *Ann Am Thorac Soc* 2014;11(8):1176–85.
- [45] Pezoulas VC, et al. A computational workflow for the detection of candidate diagnostic biomarkers of Kawasaki disease using time-series gene expression data. *Comput Struct Biotechnol J* 2021;19:3058–68.
- [46] Rosas IO, et al. MMP1 and MMP7 as potential peripheral blood biomarkers in idiopathic pulmonary fibrosis. *PLOS Med* 2008;5(4):e93.
- [47] Adams TS, et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv* 2020;6(28):eaba1983.
- [48] Habermann AC, et al. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci Adv* 2020;6(28):eaba1972.
- [49] Tsitoura E, et al. Collagen 1a1 expression by airway macrophages increases in fibrotic ILDs and is associated with FVC decline and increased mortality. *Front Immunol* 2021;12:4548.
- [50] Tsukui T, et al. Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nat Commun* 2020;11(1):1920.
- [51] Vuga LJ, et al. Cartilage oligomeric matrix protein in idiopathic pulmonary fibrosis. *PLOS ONE* 2013;8(12):e83120.
- [52] Yamaguchi K, et al. AGER gene polymorphisms and soluble receptor for advanced glycation end product in patients with idiopathic pulmonary fibrosis. *Respirology* 2017;22(5):965–71.
- [53] Machahua C, et al. Serum AGE/RAGEs as potential biomarker in idiopathic pulmonary fibrosis. *Respir Res* 2018;19(1):215.
- [54] Fichtner-Feigl S, et al. IL-13 signaling through the IL-13 α 2 receptor is involved in induction of TGF- β 1 production and fibrosis. *Nat Med* 2006;12(1):99–106.
- [55] Chen L, et al. Identification of metastasis-associated genes in cutaneous squamous cell carcinoma based on bioinformatics analysis and experimental validation. *Adv Ther* 2022;39(10):4594–612.
- [56] Odani H, et al. Suppression of renal α -dicarbonyl compounds generated following ureteral obstruction by kidney-specific α -dicarbonyl/l-xylulose reductase. *Ann N Y Acad Sci* 2008;1126(1):320–4.
- [57] Xu B, et al. Gasdermin D plays a key role as a pyroptosis executor of non-alcoholic steatohepatitis in humans and mice. *J Hepatol* 2018;68(4):773–82.
- [58] Tabib T, et al. Myofibroblast transcriptome indicates SFRP2hi fibroblast progenitors in systemic sclerosis skin. *Nat Commun* 2021;12(1):4384.
- [59] Montagner M, et al. Crosstalk with lung epithelial cells regulates Sfrp2-mediated latency in breast cancer dissemination. *Nat Cell Biol* 2020;22(3):289–96.
- [60] Mayr CH, et al. Integrative analysis of cell state changes in lung fibrosis with peripheral protein biomarkers. *EMBO Mol Med* 2021;13(4):e12871.
- [61] Sutton RM, et al. Rare surfactant-related variants in familial and sporadic pulmonary fibrosis. *Hum Mutat*; 2022. n/a(n/a).
- [62] Kumata S, et al. Living donor lobar lung transplant for a patient with lung disease caused by ABCA3 Gene mutations: a case report. *Transplant Proc* 2022.

***ENPP2* methylation in health and cancer**

Μεθυλίωση του *ENPP2* σε υγιείς και καρκινικές συνθήκες

Η αυτοταξίνη (autotaxin; ATX), κωδικοποιούμενη από το γονίδιο της 2 εξωνουκλεοτιδικής πυροφωσφατάσης/φωσφοδιεστεράσης (*ENPP2*), είναι ένα ένζυμο κλειδί για την σχετική με τον καρκίνο σύνθεση λυσοφωσφατιδικού οξέος (LPA). Παρότι έχει ήδη αναφερθεί παρεκκλίνουσα έκφραση του γονιδίου, το προφίλ μεθυλίωσής του σε υγιείς και κακοήθεις συνθήκες δεν έχει περιγραφεί. Εξετάσαμε *in silico* την μεθυλίωση του, αναλύοντας δημοσίως διαθέσιμα συνολα δεδομένων μεθυλίωσης προκειμένου να ταυτοποιήσουμε διαφορικά μεθυλιωμένες θέσεις CpGs (DMCs). Ακολούθως οι τελευταίες συσχετίστηκαν με τιμές έκφρασης σε γονιδιακό και επίπεδο μεταγραφωμικών ισομορφών. Υγιείς ιστοί παρουσίασαν μεθυλίωση σε όλο το σώμα του γονιδίου και χαμηλότερα επίπεδα σε περιοχές συσχετιζόμενες με τον υποκινητή του, ενώ το αντίστροφο προφίλ παρατηρήθηκε στην πλειονότητα των εξετασθέντων όγκων. Τα DMCs που ταυτοποιήθηκαν στον υποκινητή τοποθετούνταν σε θέσεις που αναγνωρίζονται από πολλαπλούς μεταγραφικούς παράγοντες, προτείνοντας συμμετοχή στη γονιδιακή έκφραση. Αλλαγές στην μεθυλίωση συσχετίστηκαν με έναν πιο επιθετικό φαινότυπο καρκινικών κυτταρικών σειρών. Σε προστατικά και πνευμονικά αδενοκαρκινώματα αυξημένη μεθυλίωση των CGs του υποκίνητη συσχετίστηκε με μειωμένη έκφραση του *ENPP2* και δυσοίωνες προγνωστικές παραμέτρους. Συγκεντρωτικώς, τα αποτελέσματά μας υποστηρίζουν πως η μεθυλίωση είναι ένα ενεργό επίπεδο ρύθμισης της έκφρασης της ATX στον καρκίνο. Η μελέτη μας, παρέχει μία εκτενή περιγραφή της κατάστασης μεθυλίωσης του *ENPP2* σε υγιείς και καρκινικές συνθήκες και επισημαίνει συγκεκριμένα DCMs ως πιθανούς προγνωστικούς βιοδείκτες.



Article

ENPP2 Methylation in Health and Cancer

Maria Panagopoulou ¹, Dionysios Fanidis ² , Vassilis Aidinis ² and Ekaterini Chatzaki ^{1,3,*}

¹ Laboratory of Pharmacology, Medical School, Democritus University of Thrace, GR-68100 Alexandroupolis, Greece; mpanagop@med.duth.gr

² Biomedical Sciences Research Center Alexander Fleming, Institute of Bio-Innovation, GR-16672 Athens, Greece; fanidis@fleming.gr (D.F.); aidinis@fleming.gr (V.A.)

³ Institute of Agri-Food and Life Sciences, Hellenic Mediterranean University Research Centre, GR-71410 Heraklion, Greece

* Correspondence: achatzak@med.duth.gr; Tel./Fax: +30-255-103-0533

† Current address: Ekaterini Chatzaki, Medical School, Laboratory of Pharmacology, Office 35, Bldg 5B, Dragana, 68100 Alexandroupolis, Greece.

Abstract: Autotaxin (ATX) encoded by Ectonucleotide Pyrophosphatase/Phosphodiesterase 2 (*ENPP2*) is a key enzyme in Lysophosphatidic Acid (LPA) synthesis implicated in cancer. Although its aberrant expression has been reported, *ENPP2* methylation profiles in health and malignancy are not described. We examined in silico the methylation of *ENPP2* analyzing publicly available methylome datasets, to identify Differentially Methylated CpGs (DMCs) which were then correlated with expression at gene and isoform levels. Significance indication was set to be FDR corrected *p*-value < 0.05. Healthy tissues presented methylation in all gene body CGs and lower levels in Promoter Associated (PA) regions, whereas in the majority of the tumors examined (HCC, melanoma, CRC, LC and PC) the methylation pattern was reversed. DMCs identified in the promoter were located in sites recognized by multiple transcription factors, suggesting involvement in gene expression. Alterations in methylation were correlated to an aggressive phenotype in cancer cell lines. In prostate and lung adenocarcinomas, increased methylation of PA CGs was correlated to decreased *ENPP2* mRNA expression and to poor prognosis parameters. Collectively, our results corroborate that methylation is an active level of ATX expression regulation in cancer. Our study provides an extended description of the methylation status of *ENPP2* in health and cancer and points out specific DMCs of value as prognostic biomarkers.

Keywords: autotaxin; *ENPP2*; methylation; cancer; expression; regulation



Citation: Panagopoulou, M.; Fanidis, D.; Aidinis, V.; Chatzaki, E. *ENPP2* Methylation in Health and Cancer. *Int. J. Mol. Sci.* **2021**, *22*, 11958. <https://doi.org/10.3390/ijms222111958>

Academic Editor: Agustín F. Fernandez

Received: 29 September 2021
Accepted: 1 November 2021
Published: 4 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

ATX encoded by *ENPP2* is a secreted lysophospholipase D (lysoPLD) and belongs to the ENPP (1–7) protein family [1]. ATX is responsible for the synthesis of the majority of extracellular LPA in blood. LPA acts locally upon increased ATX levels through at least six G protein-coupled receptors [2]. Increased ATX activity and levels have been correlated with several inflammatory [3] and fibroproliferative conditions [4], as well as with cancer [5]. In particular, increased expression of ATX in blood and the subsequent increase of LPA have been correlated with cancer invasiveness [6]. In addition, it has been shown that ATX expression is upregulated in cancerous [7,8] and fibrotic tissues [9].

ENPP2 contains 26 introns and 27 exons and is located in the human chromosomal region 8q24 [10], a region with frequent genetic alterations in many cancers [11]. *ENPP2* is characterized by alternative splicing of mRNA. The best-known splice variants of *ENPP2* are isoforms alpha, beta and gamma; between them, differences in the stability and expression pattern have been documented among several tissues [12].

Epigenetic regulation of *ENPP2* has been previously reported [13]. DNA methylation, a well-studied epigenetic mechanism, can regulate gene expression [14], and aberrant gene-specific methylation has been correlated with many pathologies, such as cancer [15–20]. However, data on the methylation profile of *ENPP2* in health and pathology are fragmented.

A study in Biliary Atresia (BA) showed hypomethylation at four CpGs of *ENPP2* promoter in the blood and liver of BA patients in relation to normal tissue and was correlated to increased ATX expression [21]. *ENPP2* promoter hypermethylation and gene under-expression was found in primary invasive breast carcinomas [13]. Similarly, in breast cancer cell lines, a promoter-associated CpG (cg02156680) of *ENPP2* was found highly methylated [22].

In the present study, we studied *in silico* the methylation of *ENPP2* in health and several malignancies and correlated it with gene and isoform expression, aggressiveness and prognosis. Cancer types included in our study were chosen based on their high world incidence, mortality and prevalence [23], as well as access to readily available suitable high-throughput datasets. We examined publicly available methylation datasets from readings by the Illumina methylation bead-chip arrays found in *Gene Expression Omnibus* (GEO), to identify Differentially Methylated CpGs (DMCs) of *ENPP2* between health and disease. Lung, prostate and liver cancer presented a greater number of Promoter Associated (PA) DMCs for *ENPP2* and were further pursued using large datasets retrieved from The Cancer Genome Atlas (TCGA), which allowed DMC correlation to clinicopathological parameters and gene expression. A workflow of our study is presented in Figure 1.

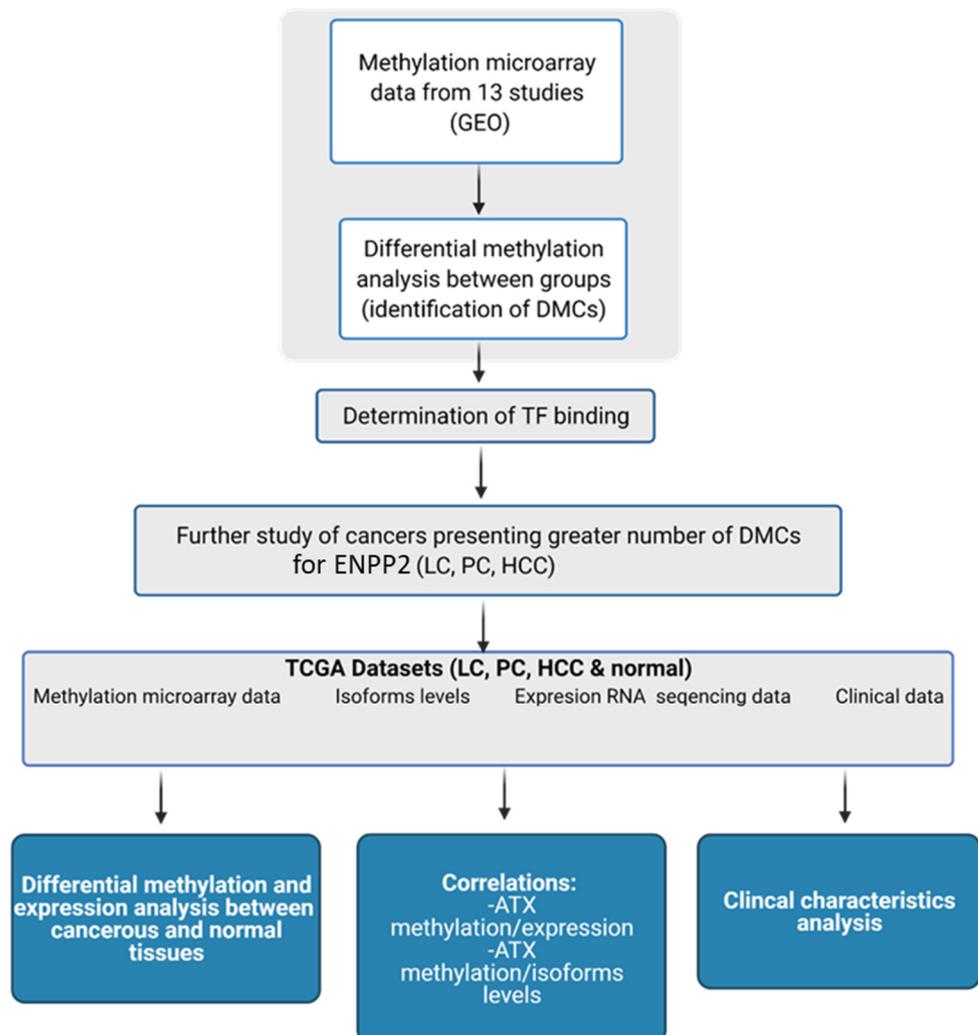


Figure 1. Workflow of the study of *ENPP2*. Created with BioRender.com (Agreement number: UW22ZTY5U7) (accessed on 24 September 2021). Abbreviations: GEO: Gene Expression Omnibus, DMCs: Differentially Methylated CpG sites, TF: Transcription Factor, TCGA: The Cancer Genome Atlas, LC: Lung Cancer, PC: Prostate Cancer, HCC: Hepatocellular Carcinoma, ATX: Autotaxin.

2. Materials and Methods

2.1. GEO Datasets for the In-Silico Methylation Analysis

DNA methylation data from cancer cell lines and patients with different malignancies and corresponding controls were obtained from the GEO (Gene Expression Omnibus) (<https://www.ncbi.nlm.nih.gov/geo/>, accessed on 25 December 2020) database [24]. Colorectal Cancer (CRC), Lung Cancer (LC), melanoma, Prostate Cancer (PC), Gastric Cancer (GC), Liver Cancer (HCC), cell lines and normal tissues were used as keywords in GEO query and ‘Methylation profiling by array’ as study type. A total of 73 studies were found; between them, only those using Infinium Human Methylation 27 K, 450 K and EPIC BeadChips (San Diego, CA, USA) and providing adequate data were selected for further analysis. In total, 13 studies, GSE27130 [25], GSE98534 [26], GSE63704, GSE46306 [27], GSE134772 [28], GSE120878 [29], GSE76938 [30], GSE97686 [31], GSE113017 [32], GSE113019 [32], GSE71627 [33], GSE50192 [34,35] and GSE51815 [36], were recruited for our analysis. Descriptions of study groups and correlations are presented in Table 1.

Table 1. Methyloome datasets retrieved from GEO for the in silico analysis of *ENPP2* methylation.

Dataset	Platform	Compared Patient Groups	References
GSE27130	27 k	236 CRC vs. 236 adjacent colon tissues	[25]
GSE98534	27 k	16 CRC vs. 16 adjacent colon tissues	[26]
GSE63704	450 k	17 LC vs. 43 adjacent lung tissues	-
GSE46306	450 k	6 CC (HPV ⁺) vs. 18 CIN3(HPV ⁺) vs. 20 normal cervical (HPV ⁻) tissues	[27]
GSE134772	EPIC	3 CC(HPV16) vs. 2 CIN1, 1 CIN2, 1 CIN3 (HPV16) vs. 3 normal (HPV ⁻) cervical tissues	[28]
GSE120878	450 K	89 melanoma vs. 73 nevus tissues	[29]
GSE76938	450 K	73 PC vs. 63 prostate benign tissues	[30]
GSE97686	450 k	3 GC vs. 3 adjacent gastric vs. 3 normal gastric myofibroblasts	[31]
GSE113017	450 k	30 HCC vs. 30 adjacent liver tissues	[32]
GSE113019	450 k	19 primary HCC vs. 18 recurrent HCC vs. 18 adjacent liver tissues	[32]
GSE71627	450 k	HepG2 vs. SKHep1, LNCaP vs. PC3	[33]
GSE50192	450 k	Adipose abdominal, adipose subcutaneous, splenic artery, bone, bone marrow red, coronary artery, gastric mucosa, lymph node, aorta abdominal, aorta thoracic, bladder, bone marrow yellow, gallbladder, ischiatic nerve, joint cartilage, medulla oblongata, tonsils (basal methylation)	[34,35]
GSE51815	450 k	AZA treated colon cancer cell lines vs. untreated controls	[36]

CRC: Colorectal Cancer; LC: Lung Cancer; CC: Cervical Cancer; HPV^{+/−}: Human Papillomavirus positive/negative; CIN: Cervical Intraepithelial Neoplasia; PC: Prostate Cancer; GC: Gastric Cancer; HCC: Hepatocellular Carcinoma.

2.2. Methylation and Statistical Analysis

Methylation analysis was carried out using normalized beta values ranging between 0 (no methylation) and 1 (full methylation) representing methylation levels of each CpG site (Level 3 data). The Kolmogorov–Smirnov test was applied to check for normality in distribution. Statistical analysis was performed using IBM SPSS 19.0 statistical software (IBM Corp. 2010. IBM SPSS Statistics for Windows, Version 19.0. Armonk, NY, USA). One-Way ANOVA tests followed by Bonferroni post-hoc or Kruskal–Wallis tests were used for comparisons of continuous variables among three or more different groups. In the case of binary variables, *t*-test or Mann–Whitney tests were also applied. Pearson or Spearman correlation was applied to compare two continuous variables. Differentially Methylated CpGs (DMCs) for *ENPP2* were identified based on the False Discovery Rate (FDR—adjusted *p*-value < 5.00 × 10^{−2}).

2.3. In Silico Determination of Transcription Factor (TF) Binding

In order to examine if the DMCs identified were correlated to *ENPP2* expression, we further analyzed promoter regions to locate TF binding sites. Hence, PROMO (<http://algen.lsi.upc.es/>, accessed on 18 March 2021) [37] tool was used in order to define possible TFs binding in identified DMCs inside *ENPP2* promoter. Only human factors and human sites were considered for a TFs search.

2.4. Expression and Methylation Correlation Analysis Using TCGA Datasets

Normalized (gene and isoform level) RNA-seq (Illumina HiSeq), level 3 methylation legacy data (Infinium Human Methylation 450 K bead-chip) and corresponding available clinical data were retrieved from prostate adenocarcinoma, lung adenocarcinoma and liver hepatocellular carcinoma TCGA projects representing PC, LC and HCC cases, using the TCGAbiolinks R package [38]. In total, 212 LC, 235 PC and 241 HCC cancer cases along with adjacent healthy lung (15 cases), prostate (35 cases) and liver (42 cases) tissues were obtained. More specifically, cases were chosen to include both mRNA expression (gene and isoforms) and methylation data, from which all matched control and tumor cases were retrieved along with 200 additional tumor samples per cancer type. In the rare case of a case ID being represented by more than one methylation or expression file, the weighted average of the respective values was used for downstream analysis (all weights sum up to the unit). Spearman correlation was calculated per cancer type using the `cor.test` function. The cutoff level of significance was set to be FDR corrected p -value < 0.05 . Last, p -values of linear models fitted between methylation and expression levels (`lm` R function) were used to test and establish the importance of small correlation coefficients. Analyses were performed using R version 4.0.4.

2.5. Expression, Methylation and Survival Analysis Using the UALCAN Database

In order to further verify our results, we used the UALCAN database (<http://ualcan.path.uab.edu/>, accessed on 10 September 2021) [39] that enables researchers to analyze cancer archived omics data. We performed expression, methylation and survival analysis of *ENPP2* in the three types of cancer used in our analysis (PC, LC, HCC) along with the corresponding controls. According to UALCAN, different beta value cut-offs have been considered to indicate hyper-methylation [beta value: 0.7 – 0.5] or hypo-methylation [beta-value: 0.3 – 0.25]. For mRNA expression, methylation and survival, we used TCGA gene analysis, and the screening conditions were as follows: gene “*ENPP2*”, TCGA dataset “Prostate adenocarcinoma”, “Lung adenocarcinoma”, “Lung Squamous cell carcinoma”, “Liver hepatocellular carcinoma”. We then used “expression”, “methylation” and “survival” as links for analysis of each cancer. Protein expression data were available only for lung adenocarcinoma, and Clinical Proteomic Tumor Analysis Consortium (CPTAC) datasets were used for analysis. For protein analysis, Z-values represent standard deviations from the median across samples for the given cancer type. Log₂ Spectral count ratio values from CPTAC were first normalized within each sample profile and then normalized across samples.

3. Results

3.1. Analysis of *ENPP2* Methylation from GEO Datasets

In silico methylation analysis of *ENPP2* was performed using methylome data retrieved using the GEO. The results are described below.

3.2. *ENPP2* Methylation in Normal Tissues

In order to examine the methylation profile of *ENPP2* across different human healthy tissues, we analyzed methylome data from 17 healthy tissues included in the GSE50192 study. We observed a consistent methylation pattern across all studied tissues (Figure 2), with methylation being increased in all 7 CGs in the gene body region and decreased in 5 CGs in the Transcription Start Site (TSS) and 1 in the 1st exon.

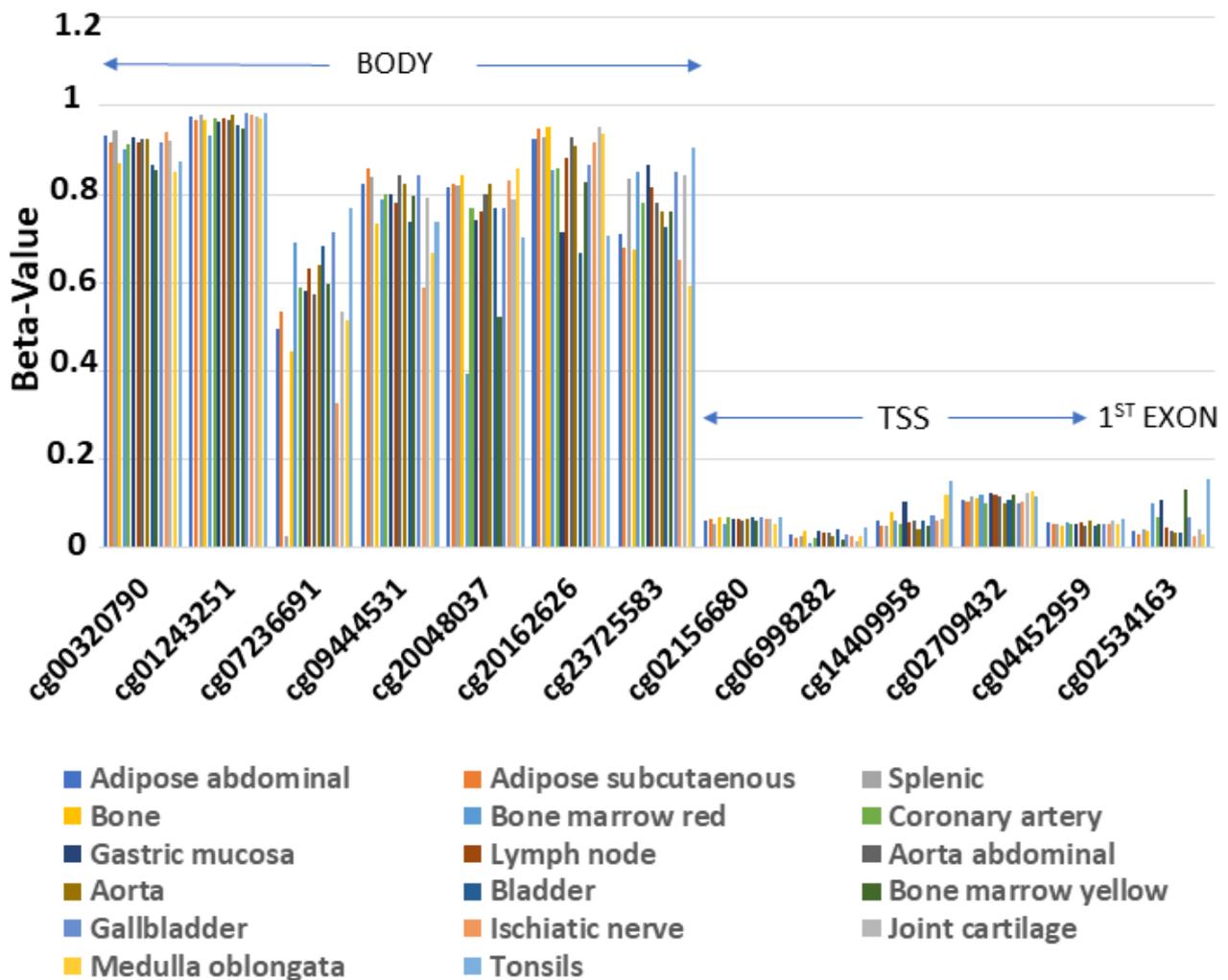


Figure 2. Similar methylation pattern of *ENPP2* across 17 healthy tissues. Gene body region presents higher methylation levels in relation to TSS or 1st exon. Abbreviations: TSS = Transcription Start Site.

3.3. *ENPP2* Methylation in Tissues from Cancer Patients

In order to unravel aberrant *ENPP2* methylation in cancer, we compared methylomes of malignant vs. respective benign tissues from 7 different cancer types, using 10 GEO datasets (GSE113017, GSE113019, GSE120878, GSE27130, GSE63704, GSE76938, GSE98534, GSE46306, GSE134772, GSE97686) (Table 1). In total, 13 DMCs were identified in 5 cancers, i.e., HCC (12 DMCs), PC (10 DMCs), LC (9 DMCs), melanoma (7 DMCs), CRC (1 DMC) (Table 2), most of which were common between them (Table 3), whereas no DMCs were identified in Precancerous Interepithelial Cervical Neoplasia (CIN) and Cancer (CC) and in Gastric Cancer (GC). With two exceptions, all gene body DMCs showed decreased methylation in cancer in relation to their controls. Most importantly, all CGs located in the promoter-associated region and the 1st exon, regions known to hold an important role in transcriptional regulation [38,40] were DMCs across different cancer types, all presenting increased methylation. These results demonstrate aberrant methylation of *ENPP2* in the majority of cancer types studied, following a specific pattern pointing to down-regulation of expression.

Table 2. DMCs presenting higher or lower methylation in cancer in relation to corresponding control, identified via in silico analysis in *ENPP2*.

Study ID	Compared Groups	CG ID	M β Value 1 *	M β Value 2 *	$\Delta\beta$ Value #	Regulation	Gene Region	FDR	
HCC									
GSE113017	Adjacent liver vs. HCC tumor	cg00320790	0.892	0.761	-0.130	Down	Body	2.5×10^{-2}	
		cg07236691	0.832	0.659	-0.173	Down	Body	1.9×10^{-3}	
		cg09444531	0.795	0.619	-0.176	Down	Body	8.6×10^{-4}	
		cg20162626	0.809	0.642	-0.167	Down	Body	3.0×10^{-2}	
		cg02709432	0.123	0.263	0.141	Up	TSS200	1.9×10^{-2}	
		cg04452959	0.058	0.189	0.131	Up	TSS200	2.1×10^{-2}	
		cg02156680	0.067	0.202	0.135	Up	TSS1500	4.0×10^{-3}	
		cg06998282	0.086	0.286	0.200	Up	TSS1500	4.0×10^{-3}	
GSE113019	Adjacent liver vs. primary HCC tumor	cg00320790	0.835	0.623	-0.212	Down	Body	3.0×10^{-2}	
		cg07236691	0.776	0.563	-0.213	Down	Body	4.5×10^{-2}	
		cg09444531	0.721	0.484	-0.237	Down	Body	2.4×10^{-3}	
		cg20048037	0.696	0.474	-0.222	Down	Body	2.1×10^{-2}	
		cg20162626	0.695	0.447	-0.248	Down	Body	7.0×10^{-3}	
		cg23725583	0.718	0.557	-0.161	Down	Body	6.2×10^{-2}	
		cg02156680	0.064	0.188	0.124	Up	TSS1500	1.8×10^{-2}	
			cg06998282	0.105	0.284	0.179	Up	TSS1500	3.0×10^{-2}
			cg02709432	0.127	0.256	0.129	Up	TSS200	2.0×10^{-2}
			cg04452959	0.042	0.149	0.106	Up	TSS200	9.3×10^{-3}
		Adjacent liver vs. (primary & recurrent) HCC tumor	cg00320790	0.835	0.634	-0.201	Down	Body	2.3×10^{-3}
			cg07236691	0.776	0.568	-0.207	Down	Body	2.3×10^{-3}
			cg20048037	0.696	0.450	-0.246	Down	Body	1.8×10^{-3}
			cg23725583	0.718	0.581	-0.137	Down	Body	4.6×10^{-2}
	cg02156680		0.064	0.198	0.134	Up	TSS1500	4.0×10^{-3}	
	cg06998282		0.105	0.279	0.174	Up	TSS1500	1.3×10^{-2}	
	cg14409958		0.263	0.374	0.111	Up	TSS1500	2.2×10^{-2}	
		cg02709432	0.127	0.253	0.126	Up	TSS200	1.2×10^{-2}	
		cg04452959	0.042	0.141	0.098	Up	TSS200	5.3×10^{-3}	
		cg02534163	0.143	0.274	0.131	Up	1st Exon	2.3×10^{-3}	
Melanoma									
GSE120878	Nevus vs. primary melanoma tissues	cg23725583	0.481	0.583	0.102	Up	Body	4.7×10^{-6}	
		cg00320790	0.859	0.834	-0.025	Down	Body	2.3×10^{-3}	
		cg09444531	0.736	0.692	-0.043	Down	Body	1.1×10^{-2}	
		cg20048037	0.801	0.740	-0.061	Down	Body	1.4×10^{-6}	
		cg20162626	0.625	0.575	-0.050	Down	Body	2.3×10^{-3}	
		cg26078665	0.647	0.616	-0.031	Down	Body	8.0×10^{-3}	
		cg04452959	0.105	0.142	0.037	Up	TSS200	5.6×10^{-4}	
		cg02534163	0.155	0.242	0.087	Up	1st Exon	1.4×10^{-6}	
CRC									
GSE27130	Adjacent colon vs. CRC	cg14409958	0.201	0.210	0.009	Up	TSS1500	5.6×10^{-3}	

Table 2. Cont.

Study ID	Compared Groups	CG ID	Mβ Value 1 *	Mβ Value 2 *	Δβ Value #	Regulation	Gene Region	FDR	
LC									
GSE63704	Normal lung vs. LC	cg09444531	0.832	0.774	−0.058	Down	Body	3.0×10^{-4}	
		cg20048037	0.832	0.754	−0.078	Down	Body	5.2×10^{-5}	
		cg20162626	0.863	0.802	−0.061	Down	Body	6.5×10^{-5}	
		cg02709432	0.234	0.265	0.030	Up	TSS200	1.7×10^{-2}	
		cg06998282	0.198	0.268	0.070	Up	TSS1500	2.6×10^{-3}	
		cg14409958	0.264	0.366	0.103	Up	TSS1500	1.4×10^{-4}	
		cg02534163	0.133	0.183	0.050	Up	1st Exon	1.5×10^{-2}	
	IPF vs. LC	cg00320790	0.940	0.904	−0.036	Down	Body	5.6×10^{-4}	
		cg20048037	0.785	0.754	−0.031	Down	Body	5.2×10^{-1}	
		cg20162626	0.844	0.802	−0.042	Down	Body	2.1×10^{-2}	
		cg06998282	0.196	0.268	0.072	Up	TSS1500	4.7×10^{-3}	
		cg14409958	0.269	0.366	0.097	Up	TSS1500	8.4×10^{-4}	
	COPD vs. LC	cg20048037	0.810	0.754	−0.056	Down	Body	8.4×10^{-3}	
		cg20162626	0.850	0.802	−0.048	Down	Body	4.7×10^{-3}	
		cg02709432	0.229	0.265	0.035	Up	TSS200	8.4×10^{-3}	
		cg02156680	0.183	0.204	0.022	Up	TSS1500	3.4×10^{-2}	
		cg06998282	0.184	0.268	0.084	Up	TSS1500	8.4×10^{-4}	
		cg14409958	0.253	0.366	0.114	Up	TSS1500	8.4×10^{-4}	
		cg02534163	0.133	0.183	0.050	Up	1st Exon	2.8×10^{-2}	
	PC								
	GSE76938	benign prostate vs. PC	cg07236691	0.471	0.670	0.199	Up	Body	1.4×10^{-6}
cg09444531			0.654	0.775	0.122	Up	Body	1.4×10^{-6}	
cg23725583			0.845	0.914	0.069	Up	Body	1.0×10^{-3}	
cg26078665			0.687	0.709	0.022	Up	Body	7.6×10^{-3}	
cg20162626			0.787	0.687	−0.100	Down	Body	1.4×10^{-6}	
cg02709432			0.093	0.417	0.324	Up	TSS200	1.4×10^{-6}	
cg04452959			0.034	0.289	0.255	Up	TSS200	1.4×10^{-6}	
cg06998282			0.108	0.457	0.349	Up	TSS1500	1.4×10^{-6}	
cg14409958			0.126	0.399	0.273	Up	TSS1500	1.4×10^{-6}	
cg02156680			0.070	0.351	0.281	Up	TSS1500	1.4×10^{-6}	
cg02534163	0.071	0.340	0.268	Up	1st Exon	1.4×10^{-6}			

* Mean β (Mβ) value 1 represents methylation in control and Mean β (Mβ) value 2 methylation in cancer; # Δβ value: (Mean β value 2-Mean β value 1). Abbreviations: DMC: Differentially Methylated CpG; FDR: False Discovery Rate; HCC: Hepatocellular Carcinoma; TSS: Transcription Start Site; PA: Promoter Associated; CRC: Colorectal Cancer; IPF: Idiopathic Pulmonary Fibrosis; LC: Lung Cancer; COPD: Chronic Obstructive Pulmonary Disease; PC: Prostate Cancer.

Table 3. Common DMCs across different cancer types, located at TSS or 1st exon, all presenting upregulation of methylation in relation to respective benign controls.

DMC	Studies Analysed	Cancer Type
cg02156680	GSE113017, GSE113019, GSE63704, GSE76938	HCC, PC
cg02709432	GSE113017, GSE113019, GSE63704, GSE76938	HCC, LC, PC
cg04452959	GSE113017, GSE113019, GSE120878, GSE76938	HCC, melanoma, PC
cg06998282	GSE113017, GSE113019, GSE63704, GSE76938	HCC, LC, PC
cg02534163	GSE113019, GSE120878, GSE63704, GSE76938	HCC, melanoma, LC, PC
cg14409958	GSE113019, GSE27130, GSE63704, GSE76938	HCC, CRC, PC, LC

Abbreviations: CRC: Colorectal Cancer; LC: Lung Cancer; PC: Prostate Cancer; HCC: Hepatocellular Carcinoma.

3.4. ENPP2 Methylation Was Correlated to Aggressiveness in Cancer Cell Lines

In order to study any relation of *ENPP2* methylation to cancer aggressiveness, we compared cell lines from hepatocellular and prostate cancer presenting a more (SKHEP1 and PC3 respectively) or less (HEPG2 and LNCAP respectively) invasive phenotype, using the GSE71627 study dataset [33]. In total, 12 DMCs were identified (Table 4), 6 of which were common in both cancer types. All 8 DMCs identified in HCC cell lines were also found in the liver tumor methylomes, whereas the common DMCs between prostate tissues and cell lines were 6/8. With two exceptions, all DMCs across the whole gene (1st exon, TSS, body) showed higher methylation in the more aggressive hepatocellular and prostate cell lines. These observations suggest an involvement of *ENPP2* methylation in cancer aggressiveness.

Table 4. DMCs of *ENPP2* identified by comparing HCC and PC cell lines with a more (SKHEP1 and PC3 respectively) to less (HEPG2 and LNCAP respectively) invasive phenotype (GSE71627 dataset).

CG ID	M β Value 1 *	M β Value 2 *	$\Delta\beta$ Value #	Regulation	Gene Region	FDR
HCC						
cg00320790	0.808	0.924	0.115	Up	Body	3.1×10^{-2}
cg09444531	0.353	0.765	0.411	Up	Body	6.5×10^{-3}
cg20048037	0.564	0.859	0.295	Up	Body	2.4×10^{-2}
cg07236691	0.689	0.126	-0.563	Down	Body	3.6×10^{-3}
cg04452959	0.339	0.784	0.444	Up	TSS200	5.1×10^{-2}
cg02156680	0.472	0.868	0.396	Up	TSS1500	3.2×10^{-3}
cg06998282	0.637	0.938	0.302	Up	TSS1500	1.8×10^{-4}
cg02534163	0.710	0.967	0.257	Up	1st Exon	2.7×10^{-2}
PC						
cg00320790	0.574	0.900	0.326	Up	Body	2.3×10^{-3}
cg07236691	0.778	0.824	0.046	Up	Body	5.6×10^{-2}
cg09444531	0.251	0.585	0.333	Up	Body	4.3×10^{-3}
cg20048037	0.370	0.671	0.301	Up	Body	8.8×10^{-3}
cg20162626	0.190	0.552	0.362	Up	Body	1.7×10^{-2}
cg26078665	0.662	0.772	0.110	Up	Body	2.3×10^{-3}
cg02156680	0.725	0.377	-0.348	Down	TSS1500	2.5×10^{-3}
cg02534163	0.772	0.951	0.179	Up	1st Exon	2.9×10^{-3}

* Mean β (M β) value 1 represents methylation in less invasive cell lines and * Mean β (M β) value 2 methylation in more invasive; # $\Delta\beta$ value: (Mean β value 2-Mean β value 1). Abbreviations: PC: Prostate Cancer, HCC: Hepatocellular Carcinoma, PA: Promoter Associated, TSS: Transcription Start Site.

Interestingly, treatment of colon cancer cell lines with the DNA methylation inhibitor 5-aza-2'-deoxycytidine (GSE51815 study) caused a decrease of methylation in all 14 DMCs located throughout *ENPP2* (Supplementary Table S1), implying that methylation could present a potential therapeutic target to reverse the aggressive phenotype.

3.5. In Silico Analysis of TF Binding on the ENPP2 Promoter

Regulation of gene expression via DNA methylation occurs mainly by disturbing TF and RNA polymerase binding to sites known to be necessary for initiation of transcription [41]. To support that the identified DMCs on *ENPP2* may actually play a role in regulating ATX expression, we examined if they are located within TF binding promoter regions that could initiate transcription. Analysis using the PROMO tool predicted 39 putative TFs that could bind to the *ENPP2* promoter (Figure 3), 4 of them (TFIID, GR, GR-beta, C/EBPbeta) on or in proximity to cg04452959, 7 TFs (TFII-I, GR-alpha, GATA-1, E2F-1, Pax-5, p53, Sp1) in cg02709432, 7 TFs (C/EBPbeta, C/EBPalpha, Pax-5, p53, ENKTF-1, YY1, GR-beta) in cg02156680 and 3 TFs (PEA3, GATA-1, XBP-1) in cg06998282. Interestingly, those 4 CGs located 200 nucleotides upstream of and up to the TSS (TSS200) (first 2), or 200 to 1500 nucleotides upstream of the TSS (TSS1500) (last 2) were identified as DMCs in most of the malignancies examined and between more and less aggressive cell lines. Collectively, these findings show that DMCs identified in the *ENPP2* promoter in cancer are found in sites significant for TF binding, and therefore, altered methylation is likely to affect transcription and expression of ATX.



Figure 3. 39 Transcription factors that can bind to the *ENPP2* promoter in sites containing DMSs identified in cancer, with a dissimilarity margin $\leq 15\%$.

3.6. ENPP2 Methylation and Expression Analysis from TCGA Datasets

An important objective of our study was to address if and how aberrant methylation of *ENPP2* is related to alterations in gene expression. Based on our findings, three cancer entities presenting the highest number of DMCs were selected for further study, in order to confirm altered *ENPP2* methylation in larger cohorts and correlate them with expression at gene and isoform levels. For this purpose, several available datasets including PC, LC and HCC readings of mRNA expression (gene and isoforms) and methylation along with the available clinical and demographic data were downloaded from TCGA.

3.7. ENPP2 Methylation and Expression Analysis in Prostate Cancer

Comparisons were performed between 235 prostate adenocarcinoma tumors and 35 healthy prostate tissues (Table 5). In general, results confirmed those from the GEO datasets. In total, 12 DMCs were identified between cancer and control tissues (5 in TSS, 1 in 1st exon and 6 in the gene body), 10 of which were common to those found in the GSE76938 dataset. All DMCs in TSS and the 1st exon presented increased methylation in PC in relation to controls, whereas decreased methylation was noticed in 3 out of 6 DMCs in the gene body area. Following this, we examined correlation of DMCs to clinicopathological patient characteristics, to reveal associations with prognosis. Methylation analysis in relation to available patient data (age, race, nodal status, relapse, tumor size and treatment response) showed a significant correlation with tumor size, as increased methylation of 3 CGs, namely, cg02534163 (1st exon), cg02709432 (TSS200) and cg23725583 (gene body), was found in larger tumors in relation to smaller tumors ($p = 0.032$). Furthermore, non-response to pharmacotherapy was correlated with increased methylation of cg01243251 in the gene body region ($p = 0.023$). No other correlations were found in relation to age, race, nodal status and the event of relapse.

Table 5. Differential methylation and expression analysis of *ENPP2* between normal prostate and PC tumors from TCGA cases.

CG ID	Mβ Value 1 *	Mβ Value 2 *	Δβ Value #	Regulation	Gene Region	FDR
cg07236691	0.568	0.730	0.162	Up	Body	1.44×10^{-6}
cg09444531	0.697	0.735	0.038	Up	Body	1.09×10^{-3}
cg20048037	0.885	0.815	-0.070	Down	Body	4.22×10^{-3}
cg20162626	0.796	0.650	-0.146	Down	Body	1.44×10^{-6}
cg23725583	0.858	0.877	0.019	Up	Body	9.05×10^{-3}
cg01243251	0.933	0.904	-0.029	Down	Body	1.44×10^{-6}
cg14409958	0.251	0.590	0.339	Up	TSS1500	1.44×10^{-6}
cg02156680	0.156	0.543	0.386	Up	TSS1500	1.44×10^{-6}
cg06998282	0.212	0.645	0.433	Up	TSS1500	1.44×10^{-6}
cg02709432	0.213	0.586	0.372	Up	TSS200	1.44×10^{-6}
cg04452959	0.130	0.466	0.335	Up	TSS200	1.44×10^{-6}
cg02534163	0.128	0.565	0.436	Up	1st Exon	1.44×10^{-6}

Δβ value: (Mean β value 2 * cancer-Mean β value 1 * normal). Abbreviations: PC: Prostate Cancer, FC: Fold Changes.

mRNA expression analysis in the same samples showed decreased levels in PC in relation to normal tissues (LogFC: -0.379 , FDR: 3.70×10^{-2}), indicating that the increased methylation of *ENPP2* in PA regions is correlated with the decreased expression of *ENPP2* in PC. Spearman correlation of mRNA *ENPP2* expression (at gene and isoform level) per CG site revealed statistically significant correlations shown in Figure 4A and Table 6. Between gene body CGs, a tendency towards positive correlation of mRNA expression to cg01243251 and cg20162626 methylation was observed and a negative to cg07236691. TSS CG sites cg02156680, cg02709432, cg06998282, cg14409958 and 1st exon cg02534163 methylation showed a negative correlation with expression, showing that the increased methylation at these regions is associated with decreased expression. Interestingly, although Spearman's coefficient is relatively small for TSS and 1st exon CGs, successful fit of a linear model further supports the existence of an expression-methylation relationship (Table 6, coefficient *p*-value column). No significant correlations emerged between *ENPP2* expression and methylation in control prostate tissue.

Table 6. Spearman correlation coefficient between *ENPP2* CG methylation and mRNA expression ($p < 0.05$) for PC, LC and HCC samples, showing mainly a negative correlation with PA CG methylation and in most cases a positive correlation with gene body methylation.

Sample Type	CG	Gene Region	Rho	FDR	Correlation	Coefficient <i>p</i> -Value
PC						
Tumor	cg06998282	TSS1500	-0.253	1.22×10^{-3}	Negative	1.55×10^{-2}
	cg02156680	TSS1500	-0.212	2.88×10^{-3}	Negative	4.43×10^{-2}
	cg14409958	TSS1500	-0.221	2.18×10^{-3}	Negative	3.90×10^{-2}
	cg02709432	TSS200	-0.176	1.30×10^{-2}	Negative	4.39×10^{-2}
	cg02534163	1st Exon	-0.226	2.18×10^{-3}	Negative	1.30×10^{-2}
LC						
Tumor	cg06998282	TSS1500	-0.142	>0.05	Negative	3.74×10^{-3}
	cg14409958	TSS1500	-0.213	2.19×10^{-2}	Negative	1.05×10^{-4}
Control	cg07236691	Body	-0.564	>0.05	Negative	2.31×10^{-2}
HCC						
Tumor	cg00320790	Body	0.297	9.46×10^{-6}	Positive	6.29×10^{-5}
	cg01243251	Body	0.247	2.94×10^{-4}	Positive	6.90×10^{-4}
	cg07236691	Body	0.239	4.06×10^{-4}	Positive	2.28×10^{-4}
	cg09444531	Body	0.395	1.22×10^{-9}	Positive	1.53×10^{-7}
	cg20048037	Body	0.436	8.81×10^{-12}	Positive	1.02×10^{-8}
	cg20162626	Body	0.473	0.00×10^0	Positive	2.10×10^{-9}
	cg06998282	TSS1500	-0.137	>0.05	Negative	4.55×10^{-2}

Table 6. Cont.

Sample Type	CG	Gene Region	Rho	FDR	Correlation	Coefficient <i>p</i> -Value
Control	cg20162626	Body	0.422	2.63×10^{-2}	Positive	3.39×10^{-2}
	cg23725583	Body	-0.35	4.74×10^{-2}	Negative	2.61×10^{-2}
	cg02709432	TSS200	-0.361	4.56×10^{-2}	Negative	3.25×10^{-2}
	cg04452959	TSS200	-0.411	2.63×10^{-2}	Negative	2.87×10^{-2}
	cg06998282	TSS1500	-0.464	1.60×10^{-2}	Negative	4.53×10^{-3}
	cg02156680	TSS1500	-0.393	2.99×10^{-2}	Negative	3.52×10^{-2}

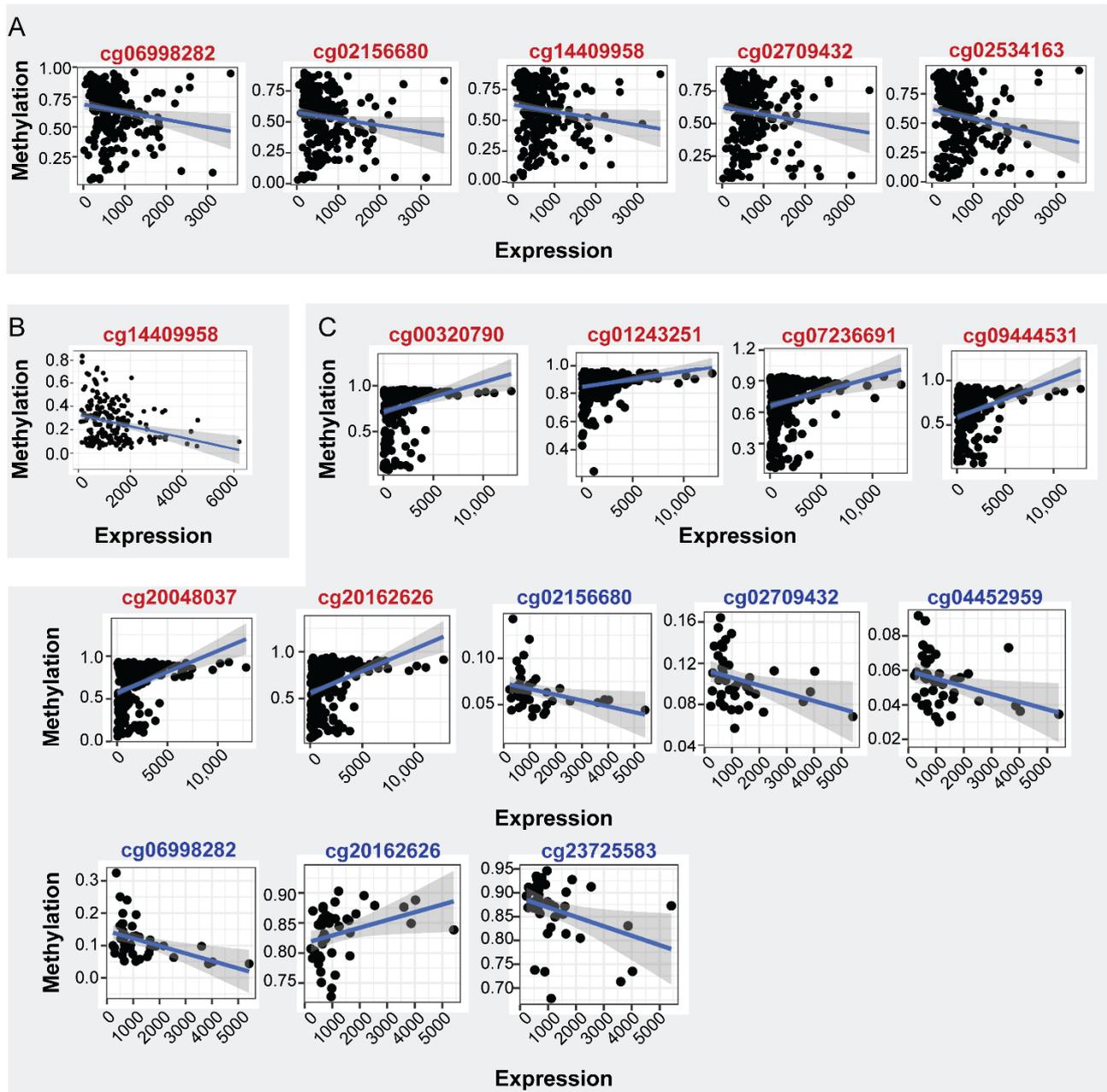


Figure 4. Correlation of ENPP2 CGs methylation and mRNA expression. CGs showing significant correlations are depicted ($p < 0.05$). (A) Expression-methylation scatter plots of CG sites of PC and normal samples, (B) expression-methylation scatter plots of CG sites of LC and normal samples, (C) expression-methylation scatter plots of CG sites of HCC and normal samples. Correlations from cancer and healthy samples are marked in red and blue CG color font, respectively.

In order to unfold the impact of CG methylation on *ENPP2* isoform expression [12,42], we downloaded mRNA expression data from *ENPP2* isoforms, i.e., isoform alpha (uc003yos.1), isoform beta (uc003yor.1 and uc003yot.1) and isoform gamma (uc010mdd.1). Uc003yot.1, uc003yos.1 and uc003yor.1 isoform expression showed statistically significant correlation with the methylation of several CGs. In specific, although they were all characterized by small effect sizes, uc003yor.1 expression is linearly related to the methylation levels of TSS and 1st exon CGs cg02156680, cg06998282, cg14409958 and cg02534163, respectively, further strengthening the observed correlation (Supplementary Figure S1 and Table S2). Finally, no relation emerged between expression of any of the *ENPP2* isoforms and methylation of CGs in healthy prostate samples, consistent with the observations at the gene level.

3.8. *ENPP2* Methylation and Expression Analysis in Lung Cancer

Analysis was performed between 212 LC adenocarcinoma tumors and 15 healthy lung tissues, and results presenting statistically significant correlations are shown in Table 7. Findings confirmed those from the GEO datasets. Eight DMCs were identified between cancer and control tissues (3 in the TSS, 1 in the 1st exon, 4 in the gene body) and 6 of them were common to those found in the GSE76938 dataset. DMCs of *ENPP2* showed upregulation of methylation in TSS (cg04452959, cg06998282, cg14409958) and the 1st exon (cg02534163) and downregulation in the gene body (cg07236691, cg09444531, cg20048037, cg20162626). Methylation was also correlated to the available clinicopathological characteristics of LC and normal lung tissue samples (gender, age, nodal status, distance metastasis, relapse, tumor size and treatment response and stage). In LC samples, increased methylation of cg14409958 (TSS) was significantly correlated with advanced cancer stage ($p = 0.035$).

Table 7. Differential methylation and expression analysis of *ENPP2* between normal lung and LC tumors from TCGA cases.

CG ID	M β Value 1 *	M β Value 2 *	$\Delta\beta$ Value #	Regulation	Gene Region	FDR
cg20162626	0.750	0.636	-0.114	Down	Body	1.19×10^{-2}
cg20048037	0.722	0.624	-0.098	Down	Body	5.18×10^{-2}
cg07236691	0.561	0.547	-0.014	Down	Body	2.95×10^{-4}
cg09444531	0.730	0.630	-0.100	Down	Body	2.95×10^{-4}
cg04452959	0.071	0.138	0.066	Up	TSS200	7.19×10^{-3}
cg14409958	0.105	0.262	0.158	Up	TSS1500	2.95×10^{-4}
cg06998282	0.096	0.230	0.134	Up	TSS1500	1.05×10^{-2}
cg02534163	0.109	0.255	0.145	Up	1st Exon	8.19×10^{-4}

M β Value 1 *: Mean β value normal, M β Value 2 *: Mean β value cancer # $\Delta\beta$ value: (Mean β value 2 * cancer - Mean β value 1 * normal).
Abbreviations: LC: Lung Cancer, FC: Fold Changes.

Differential mRNA expression analysis in the same samples showed decreased levels in LC in relation to normal tissues (LogFC: 1.285, FDR: $< 1.00 \times 10^{-2}$) similarly to PC, indicating that in cancer the increased methylation of PA CGs is correlated to decreased autotaxin expression. The impact of *ENPP2* methylation on its expression was examined in LC and healthy lung tissue samples. Spearman correlation of mRNA expression per CG resulted in a single statistically significant correlation (Figure 4B and Table 6). A reverse correlation was noticed between methylation of cg14409958 (TSS) and mRNA expression, suggesting again the DNA methylation role in repressing expression. Fit of a linear model once again reinforced the observed correlation. On the other hand, control samples did not show any statistically significant correlation after p -value correction, and only the methylation of body site cg7236691 showed a significant correlation coefficient along with a linear relationship to *ENPP2* expression levels. Last, no significant correlations were witnessed between methylation and expression levels of all isoforms examined, yet large rho values and significant linear model fit propose the existence of such a relationship.

3.9. ENPP2 Methylation and Expression Analysis in Hepatocellular Carcinoma

Analysis was performed between 241 HCC tumors and 42 control liver tissues. Statistically significant correlations are presented in Table 8. In total, 13 DMCs were identified between cancer and control (5 in the TSS, 1 in the 1st exon, 7 in the gene body) and 12 were common to those found in GSE113017 and GSE113019. Again, downregulation of methylation was noticed in all gene body CGs (cg00320790, cg23725583 and cg01243251) and upregulation of methylation in all TSS and 1st exon related CGs (cg02156680, cg02534163, cg02709432, cg04452959, cg06998282 and cg14409958) in HCC.

Table 8. Differential methylation and expression analysis of *ENPP2* between normal liver and HCC tumors from TCGA cases.

CG ID	Mβ Value 1 *	Mβ Value 2 *	Δβ Value #	Regulation	Gene Region	FDR
cg00320790	0.929	0.798	−0.131	Down	Body	3.90×10^{-5}
cg01243251	0.917	0.872	−0.046	Down	Body	1.99×10^{-3}
cg07236691	0.878	0.734	−0.143	Down	Body	1.30×10^{-5}
cg09444531	0.850	0.689	−0.161	Down	Body	1.30×10^{-5}
cg20048037	0.832	0.684	−0.148	Down	Body	4.04×10^{-2}
cg20162626	0.833	0.671	−0.161	Down	Body	3.75×10^{-4}
cg23725583	0.864	0.765	−0.099	Down	Body	1.50×10^{-2}
cg04452959	0.054	0.180	0.126	Up	TSS200	4.46×10^{-2}
cg02709432	0.104	0.264	0.160	Up	TSS200	2.83×10^{-4}
cg02156680	0.065	0.210	0.144	Up	TSS1500	3.76×10^{-2}
cg06998282	0.116	0.300	0.184	Up	TSS1500	1.86×10^{-2}
cg14409958	0.340	0.463	0.123	Up	TSS1500	1.86×10^{-2}
cg02534163	0.070	0.247	0.177	Up	1st Exon	3.15×10^{-4}

Mβ Value 1 *: Mean β value normal, Mβ Value 2 *: Mean β value cancer # Δβ value: (Mean β value cancer−Mean β value normal).
Abbreviations: LC: Lung Cancer, FC: Fold Changes.

Methylation of *ENPP2* was also correlated to available clinical and demographic characteristics of the HCC cohort. Interestingly, in the tumor samples, increased methylation of the majority of the *ENPP2* CGs (cg00320790, cg01243251, cg02156680, cg02709432, cg07236691, cg09444531, cg14409958, cg20048037, cg20162626, cg23725583) (all $p < 0.05$) was noticed in women in relation to men. In addition, a negative correlation was found between age and methylation of cg00320790, cg01243251, cg07236691, cg09444531, cg20048037, cg20162626 and cg23725583 (all $p < 0.001$), i.e., younger people presented increased methylation in relation to older. Finally, increased methylation of cg04452959 was correlated to tumors with macro invasion in relation to those with no or micro invasion ($p = 0.044$). No correlation was noticed between methylation and BMI, hepatic inflammation, Ishak fibrosis, relapse, family history, grade, stage or tumor size. Analysis in normal samples showed a gender correlation only for one CG (cg20048037) which presented increased methylation in females ($p = 0.033$) in relation to males. Negative correlation was also noticed between cg01243251 methylation and age ($p = 0.037$). Finally, no relationship was found between BMI and *ENPP2* methylation in normal samples.

mRNA expression analysis in the same samples showed increased levels in HCC in relation to normal tissues (LogFC: 0.710, FDR: 1.00×10^{-2}), i.e., the opposite of LC and PC observations, suggesting a methylation-independent and a cancer type-specific regulation of *ENPP2* in HCC. Spearman correlation of mRNA expression (at gene and isoform levels) per CG site revealed the most numerous statistically significant correlations, compared to PC and LC samples, shown in Figure 4C and Table 6. Between gene body CGs, a positive correlation of mRNA expression to cg00320790, cg01243251, cg07236691, cg09444531, cg20048037 and cg20162626 methylation was observed. Apart from the significant correlations established, methylation of the aforementioned CGs was characterized by a linear relationship to *ENPP2* expression, further supporting dependence of the latter on the former. Control samples also showed positive correlation between *ENPP2* mRNA expression and methylation of 2 gene body CGs (cg20048037, g20162626). Finally, cg23725583 of body and cg02709432, cg04452959, cg06998282 and cg02156680 of TSS regions showed negative

correlation of methylation in relation to *ENPP2* mRNA expression. Isoform analysis for the control tissues showed similar correlation patterns (Supplementary Figure S2 and Table S2).

3.10. *ENPP2* Methylation, Expression and Survival Analysis via UALCAN

In order to further verify our findings, we conducted expression, methylation and survival analysis of *ENPP2* in PC (all adenocarcinoma cases), LC (adenocarcinoma and squamous cell carcinoma cases) and HCC using the UALCAN database. Analysis confirmed the above results as *ENPP2* mRNA was under-expressed in PC ($p = 9.31 \times 10^{-3}$, Figure 5A) and LC (adenocarcinoma, $p = 1.68 \times 10^{-3}$ and squamous cell carcinoma, $p = 4.52 \times 10^{-3}$, Figure 6A,C) and upregulated in HCC ($p = 2.38 \times 10^{-10}$, Figure 7A). Protein expression analysis was available only for LC adenocarcinoma cases, showing downregulation in primary tumor tissues in relation to normal tissues (Figure 6E, $p = 1.78 \times 10^{-4}$). Next, methylation analysis revealed upregulation in all cancer types in relation to normal tissues (PC, $p = 1.62 \times 10^{-12}$, HCC, $p = 1.11 \times 10^{-16}$ and LC, $p = 1.62 \times 10^{-12}$ for both types) as depicted in Figures 5–7, in accordance with our previous observations. Methylation and expression results via UALCAN strengthen our findings, showing that the *ENPP2* gene is methylated in LC, HCC and PC and this is related to under-expression in LC and PC, suggesting a causative relationship in these two cancer types and a cancer-specific regulatory mechanism in HCC. Finally, survival analysis did not reveal any statistical significance for any of the studied cancers, as depicted in Supplementary Figure S4A–D.

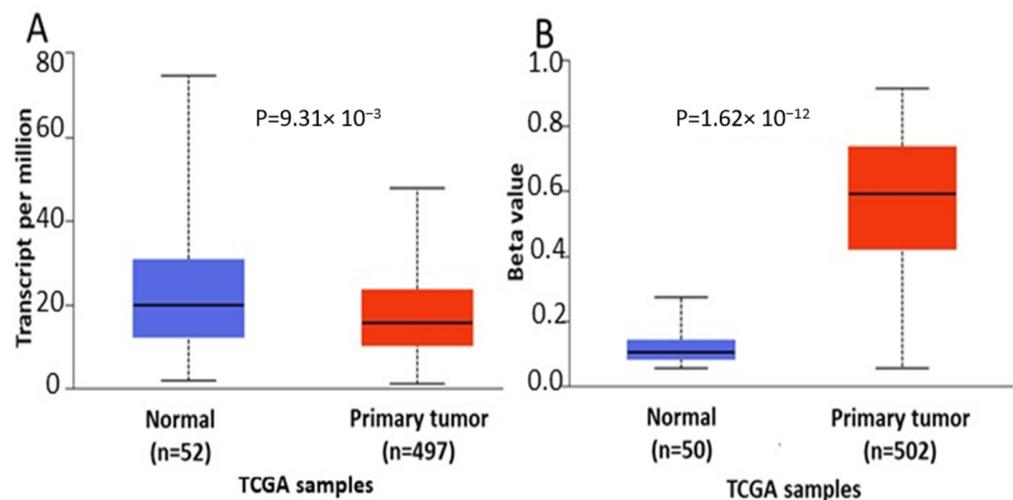


Figure 5. Analysis of *ENPP2* between primary PC tumors and normal tissues using the UALCAN database, concerning (A) mRNA expression and (B) DNA methylation. Abbreviations: TCGA: The Cancer Genome Atlas, PC: Prostate Cancer.

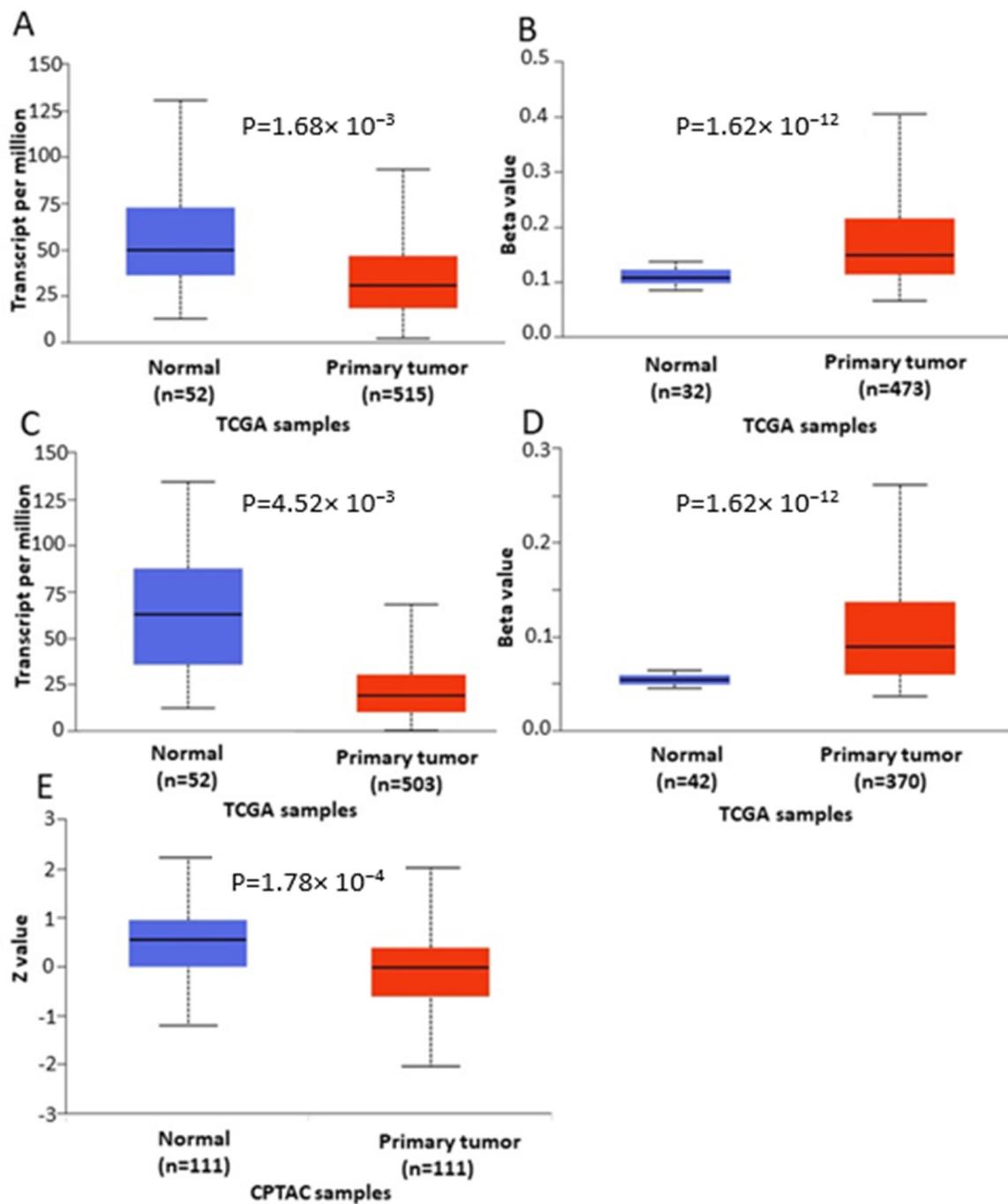


Figure 6. Analysis of *ENPP2* between primary LC tumors and normal tissues using the UALCAN database, concerning (A) mRNA expression and (B) DNA methylation of adenocarcinoma cases, (C) mRNA expression and (D) DNA methylation of squamous cell carcinoma and (E) protein expression of adenocarcinoma cases. Abbreviations: TCGA: The Cancer Genome Atlas, CPTAC: Clinical Proteomic Tumor Analysis Consortium, LC: Lung Cancer.

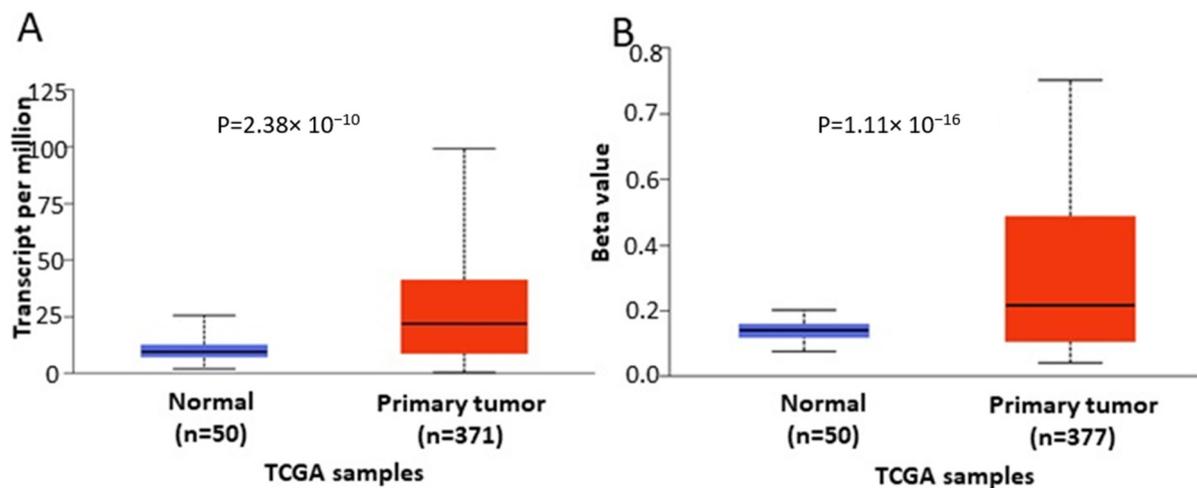


Figure 7. Analysis of *ENPP2* between primary HCC tumors and normal tissues using the UALCAN database, concerning (A) mRNA expression and (B) DNA methylation. Abbreviations: TCGA: The Cancer Genome Atlas, HCC: Hepatocellular Carcinoma.

4. Discussion

ATX encoded by *ENPP2* is a secreted glycoprotein that forms LPA [42]. The ATX-LPA axis is related to many physiological processes, including embryonic development and wound healing. Dysregulation of ATX expression is connected with various pathological conditions such as cancer, inflammatory diseases and fibrosis [3–5]. The exact mechanism by which *ENPP2* expression is regulated is still not fully understood, whereas recently, it has been proved that *ENPP2* is prone to epigenetic alterations [13]. Still, very little information is available about its DNA methylation pattern and the consequent impact in gene expression in health and human pathology.

In the present study we adopted a bioinformatic in silico approach using publicly available datasets from healthy tissues and different cancer tissues and cell lines to analyze methylation patterns of *ENPP2*. Our analysis showed a consistent methylation pattern throughout the gene's regions across human tissues, i.e., increased methylation in the gene body and decreased methylation in TSS and the 1st exon. Given the fact that *ENPP2* is expressed in almost all tissues and biological fluids [12,43,44], we can postulate that the decreased methylation in the TSS and 1st exon is associated with the active transcription of the gene in most human tissues.

Analysis of cancer datasets revealed aberrant *ENPP2* methylation, showing a malignant-specific profile throughout different cancer types. In general, methylation was increased in the TSS and 1st exon, regions known to hold an important role in gene expression, and decreased in the gene body region. A large number of DMCs were identified between malignant and respective benign tissues. Most importantly, all six DMCs of *ENPP2* located at TSS in the promoter or at the 1st exon showed increased methylation across different cancer types, including HCC, melanoma, CRC, LC and PC. These results corroborate and expand recent observations showing a hypermethylated *ENPP2* promoter in primary tumors of LC and squamous cell carcinoma patients [45] and in breast cancer [13,22,46].

Based on these interesting observations, we next performed in silico analysis of *ENPP2* methylation in datasets retrieved from the TCGA, focusing on those cancer types presenting the greatest number of DMCs, i.e., LC, PC and HCC. TCGA datasets are generally larger compared to those of other research efforts, allowing comparisons of stronger statistical relevance, and most significantly, they contain several clinical and demographic parameters of each patient. In addition, the datasets selected included also mRNA expression data and were therefore suitable for addressing an important objective of this study, i.e., if aberrant methylation is correlated to gene expression. Methylation, clinical and expression data were recovered for the three cancer types. Differential methylation analysis of *ENPP2* revealed

that all emerged DMCs identified in transcription-related (TSS and 1st exon) regions were hypermethylated in all three cancers compared to healthy controls, confirming the analysis of the GEO datasets. In addition, the majority of DMCs located at the gene body were hypomethylated in the studied cancers in relation to controls. mRNA levels were decreased in PC and LC in relation to normal tissues. Collectively, our results indicate that the increased methylation of PA and 1st exon CGs is correlated with decreased expression in lung and prostate cancer. This is in line with previous studies in LC and BC showing that *ENPP2* is hypermethylated in tumor tissues in relation to normal, causing down regulation in gene expression [13,45]. In PC, ATX protein was not or was weakly expressed in non-neoplastic epithelial cells and in high-grade intra-epithelial neoplasia, while in cancer cells ATX was only expressed in half of the tumors and was correlated with adverse tumor parameters [47]. A relevant study in LC showed that ATX protein expression and activity was increased in LC tissues and sera [48]. As far as HCC is concerned, our analysis showed upregulation of expression in HCC in relation to normal liver, showing a TSS and 1st exon methylation-independent and a cancer type-specific role of *ENPP2* expression regulation. In a previous study, ATX overexpression in HCC tissues was correlated with inflammation and liver cirrhosis. In addition, liver cancer cell lines presented stronger ATX expression in relation to normal hepatocytes [49]. It should be noted that many authors have demonstrated that the relationship between mRNA expression and protein differs in many cancers. It has been reported in lung cancer and glioblastoma that, for many genes, mRNA expression is lower but protein levels are higher compared with the control [50–53].

In agreement with the above findings, analysis using the UALCAN database showed that *ENPP2* is hypermethylated and under-expressed in LC and PC, suggesting that DNA methylation regulates expression in LC and PC. However, no regulatory relation was observed between methylation and expression in HCC, as both were upregulated, pointing again to a cancer-specific methylation-independent *ENPP2* regulation. Different mechanisms between cancer types are common. Here, our presented results from the cancer types studied indicate a cancer type-specific profile of *ENPP2* methylation rather than a similar pan-cancer dysregulation. Without availability of suitable methylome datasets or targeted methylation studies of *ENPP2* in each different cancer type, we cannot extrapolate conclusions between cancers.

The same correlation pattern was noticed for *ENPP2* isoforms in all cancer types studied. Interestingly, there was a significant negative correlation between mRNA expression (gene and isoform alpha and beta) and promoter methylation in four CGs (cg02156680, cg02709432, cg04452959 and cg06998282) in PRAD. In LC samples, the methylation of cg06998282 and cg02709432 was negatively correlated with the expression of *ENPP2* and also with isoform beta and gamma. Finally, in the case of HCC, only the methylation of cg06998282 was negatively correlated with the expression of *ENPP2* and isoform beta. The above findings indicate that the promoter methylation of specific CGs is negatively correlated with *ENPP2* and isoform expression differs between cancers, with cg02709432 being a common site in PC and LC but not in the case of HCC. This CG is located at a site that can bind E2F-1 TF, which has been shown to be inhibited by CG methylation [54], and Sp1 TF, which has been found to regulate *ENPP2* transcription [55]. Thus, we hypothesized that as the level of methylation increases, methylation of cg02709432 hinders the binding of the TFs to the promoter, thus leading to reduction in *ENPP2* gene and isoforms expression.

The expression pattern of isoforms differs between tissues as high expression levels of isoform beta were found in peripheral tissues and plasma, while isoform gamma was mostly found in the brain, and isoform alpha is considered to be the most under-expressed in brain and peripheral tissue in comparison to the other two [56]. According to a relevant study, isoform alpha has a deletion of exon 12, isoform beta a deletion of exons 12 and 21 and isoform gamma a deletion of exon 21 [12], leading to different splice variants. None of the identified DMCs were located at these regions, explaining similar patterns of *ENPP2* mRNA and isoform expression.

DNA methylation within promoters is known to modulate the binding of TFs to regulatory elements, thus resulting in transcriptional repression [57]. In our study, we predicted 39 TFs which can regulate transcription through binding to *ENPP2* promoter's DMCs. Therefore, any aberrant methylation events in these DMCs during pathological transformation may block TF binding and related transcription. This is further supported by reports involving the identified TFs in *ENPP2* and ATX expression. Indeed, among the predicted TFs, NF kappaB, AP-2 and E2F have been previously shown to be sensitive to CG methylation with consequent inhibition of their DNA binding activities [54]. Another TF predicted to bind DMCs of *ENPP2*, NFAT1, has been shown to mediate ATX overexpression in MDA-MB-435 cells [58]. It has also been shown that blocking the expression of NFAT1 results in downregulation of ATX expression, leading to inhibition of melanoma and metastasis [35]. High C-Jun levels seem to enhance *ENPP2* expression [59]. Interestingly, SP was found to regulate *ENPP2* transcription in neuroblastoma cells by activating a CRE/AP-1-like element at position –142 to –149 and a GAbbox at position –227 to –235 near the TSS of *ENPP2* [55]. This is in accordance with our finding that Sp1 can bind near the cg02709432 located at TSS200.

In order to assess any correlation of *ENPP2* methylation to tumor prognosis, clinical characteristics analysis was performed and showed that increased methylation of some CGs was correlated with poor tumor parameters. Indeed, in PC it was associated with larger tumors and non-response to pharmacotherapy, in LC it was connected to the advanced cancer stage and in HCC it was associated with macro-invasion. Hence, *ENPP2* methylation in the identified CGs could be pursued further and be evaluated in clinical cancer samples as biomarkers of cancer progression and poor outcome. In addition, these results corroborate previous data showing that low mRNA expression was associated with worse prognosis in LC [45].

The involvement of *ENPP2* methylation in tumor progression and prognosis was also addressed by analyzing methylomes from cell lines presenting a more or less aggressively invasive phenotype, revealing several DMCs. Higher methylation was observed in the more aggressive in relation to less aggressive HCC and PC cell lines, indicating a connection of *ENPP2* methylation with worse prognostic behavior, in accordance with our findings in the clinical samples.

Finally, analysis of colon cell lines treated with DNA methyltransferase inhibitors showed that 5-AZA caused a decrease of methylation in all CGs in relation to untreated controls in the three studied cell lines, showing a clear demethylation effect in the *ENPP2* gene. Given the contribution of *ENPP2* in a variety of pathologies, further studies could assess a methylation-based reprogramming of *ENPP2* via a variety of methylation inhibitors. Similarly, previous studies have demonstrated that targeting the ATX-LPA-LPP axis is an attractive strategy for introducing new therapeutic choices [60,61].

In conclusion, healthy tissues presented increased methylation of *ENPP2* in the gene body and decreased in the promoter and 1st exon connected to the active transcription of the gene in most human tissues. A different pattern was described in HCC, melanoma, CRC, LC and PC, showing a malignant-specific profile of *ENPP2* methylation. Further analysis of independent TCGA datasets confirmed these results as increased methylation of promoter and 1st exon CGs and decreased *ENPP2* mRNA expression in PC and LC in relation to healthy tissues were found. Furthermore, increased methylation of *ENPP2* was connected to poor prognostic parameters in the same cancers, which was also supported by analysis of cell line datasets. We also found a negative correlation between mRNA expression at gene and isoform levels and methylation of PA CGs that present TF binding sites. In specific, we postulate that the methylation of promoter CGs may hinder the binding of TFs, and thus, the expression of *ENPP2* and isoforms may be reduced.

Our findings contribute to the understanding of methylation events and regulatory mechanism of *ENPP2* in cancer and provide a full description of DMCs to be further validated in functional and clinical studies.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms222111958/s1>.

Author Contributions: Conceptualization M.P., V.A. and E.C.; methodology, M.P. and D.F.; software, M.P. and D.F.; validation, M.P. and E.C.; formal analysis, M.P. and D.F.; investigation, M.P., D.F., E.C. and V.A.; data curation M.P., D.F., V.A. and E.C.; writing—original draft preparation, M.P.; writing—review and editing, M.P., D.F., V.A. and E.C.; visualization, M.P. and D.F.; supervision, V.A. and E.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH—CREATE—INNOVATE (project code: T1EDK-00940).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Stefan, C.; Jansen, S.; Bollen, M. NPP-type ectophosphodiesterases: Unity in diversity. *Trends Biochem. Sci.* **2005**, *30*, 542–550. [[CrossRef](#)] [[PubMed](#)]
2. Tanaka, M.; Okudaira, S.; Kishi, Y.; Ohkawa, R.; Iseki, S.; Ota, M.; Noji, S.; Yatomi, Y.; Aoki, J.; Arai, H. Autotaxin stabilizes blood vessels and is required for embryonic vasculature by producing lysophosphatidic acid. *J. Biol. Chem.* **2006**, *281*, 25822–25830. [[CrossRef](#)]
3. Ninou, I.; Magkrioti, C.; Aidinis, V. Autotaxin in Pathophysiology and Pulmonary Fibrosis. *Front. Med.* **2018**, *5*, 180. [[CrossRef](#)] [[PubMed](#)]
4. Magkrioti, C.; Galaris, A.; Kanellopoulou, P.; Stylianaki, E.A.; Kaffe, E.; Aidinis, V. Autotaxin and chronic inflammatory diseases. *J. Autoimmun.* **2019**, *104*, 102327. [[CrossRef](#)] [[PubMed](#)]
5. Peyruchaud, O.; Saier, L.; Leblanc, R. Autotaxin Implication in Cancer Metastasis and Autoimmune Disorders: Functional Implication of Binding Autotaxin to the Cell Surface. *Cancers* **2019**, *12*, 105. [[CrossRef](#)] [[PubMed](#)]
6. Benesch, M.G.; Tang, X.; Maeda, T.; Ohhata, A.; Zhao, Y.Y.; Kok, B.P.; Dewald, J.; Hitt, M.; Curtis, J.M.; McMullen, T.P.; et al. Inhibition of autotaxin delays breast tumor growth and lung metastasis in mice. *FASEB J.* **2014**, *28*, 2655–2666. [[CrossRef](#)] [[PubMed](#)]
7. Memet, I.; Tsalkidou, E.; Tsaroucha, A.K.; Lambropoulou, M.; Chatzaki, E.; Trypsianis, G.; Schizas, D.; Pitiakoudis, M.; Simopoulos, C. Autotaxin Expression in Hepatocellular Carcinoma. *J. Investig. Surg.* **2018**, *31*, 359–365. [[CrossRef](#)] [[PubMed](#)]
8. Yang, S.Y.; Lee, J.; Park, C.G.; Kim, S.; Hong, S.; Chung, H.C.; Min, S.K.; Han, J.W.; Lee, H.W.; Lee, H.Y. Expression of autotaxin (NPP-2) is closely linked to invasiveness of breast cancer cells. *Clin. Exp. Metastasis* **2002**, *19*, 603–608. [[CrossRef](#)]
9. Oikonomou, N.; Mouratis, M.A.; Tzouveleki, A.; Kaffe, E.; Valavanis, C.; Vilaras, G.; Karameris, A.; Prestwich, G.D.; Bouros, D.; Aidinis, V. Pulmonary autotaxin expression contributes to the pathogenesis of pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* **2012**, *47*, 566–574. [[CrossRef](#)]
10. Lee, H.Y.; Murata, J.; Clair, T.; Polymeropoulos, M.H.; Torres, R.; Manrow, R.E.; Liotta, L.A.; Stracke, M.L. Cloning, chromosomal localization, and tissue expression of autotaxin from human teratocarcinoma cells. *Biochem. Biophys. Res. Commun.* **1996**, *218*, 714–719. [[CrossRef](#)]
11. Brisbin, A.G.; Asmann, Y.W.; Song, H.; Tsai, Y.Y.; Aakre, J.A.; Yang, P.; Jenkins, R.B.; Pharoah, P.; Schumacher, F.; Conti, D.V.; et al. Meta-analysis of 8q24 for seven cancers reveals a locus between NOV and ENPP2 associated with cancer development. *BMC Med. Genet.* **2011**, *12*, 156. [[CrossRef](#)]
12. Giganti, A.; Rodriguez, M.; Fould, B.; Moulharat, N.; Cogé, F.; Chomarar, P.; Galizzi, J.-P.; Valet, P.; Saulnier-Blache, J.-S.; Boutin, J.A.; et al. Murine and Human Autotaxin α , β , and γ Isoforms: Gene organization, tissue distribution, and biochemical characterization. *J. Biol. Chem.* **2008**, *283*, 7776–7789. [[CrossRef](#)]
13. Parris, T.Z.; Kovács, A.; Hajizadeh, S.; Nemes, S.; Semaan, M.; Levin, M.; Karlsson, P.; Helou, K. Frequent MYC coamplification and DNA hypomethylation of multiple genes on 8q in 8p11-p12-amplified breast carcinomas. *Oncogenesis* **2014**, *3*, e95. [[CrossRef](#)]
14. Razin, A.; Cedar, H. DNA methylation and gene expression. *Microbiol. Rev.* **1991**, *55*, 451–458. [[CrossRef](#)]
15. Nikolaidis, C.; Nena, E.; Panagopoulou, M.; Balgouranidou, I.; Karaglani, M.; Chatzaki, E.; Agorastos, T.; Constantinidis, T.C. PAX1 methylation as an auxiliary biomarker for cervical cancer screening: A meta-analysis. *Cancer Epidemiol.* **2015**, *39*, 682–686. [[CrossRef](#)] [[PubMed](#)]
16. Panagopoulou, M.; Esteller, M.; Chatzaki, E. Circulating Cell-Free DNA in Breast Cancer: Searching for Hidden Information towards Precision Medicine. *Cancers* **2021**, *13*, 728. [[CrossRef](#)]

17. Panagopoulou, M.; Karaglani, M.; Manolopoulos, V.G.; Iliopoulos, I.; Tsamardinos, I.; Chatzaki, E. Deciphering the Methylation Landscape in Breast Cancer: Diagnostic and Prognostic Biosignatures through Automated Machine Learning. *Cancers* **2021**, *13*, 1677. [[CrossRef](#)]
18. Panagopoulou, M.; Cheretaki, A.; Karaglani, M.; Balgkouranidou, I.; Biziota, E.; Amarantidis, K.; Xenidis, N.; Kakolyris, S.; Baritaki, S.; Chatzaki, E. Methylation Status of Corticotropin-Releasing Factor (CRF) Receptor Genes in Colorectal Cancer. *J. Clin. Med.* **2021**, *10*, 2680. [[CrossRef](#)] [[PubMed](#)]
19. Panagopoulou, M.; Karaglani, M.; Balgkouranidou, I.; Biziota, E.; Koukaki, T.; Karamitrousis, E.; Nena, E.; Tsamardinos, I.; Kolios, G.; Lianidou, E.; et al. Circulating cell-free DNA in breast cancer: Size profiling, levels, and methylation patterns lead to prognostic and predictive classifiers. *Oncogene* **2019**, *38*, 3387–3401. [[CrossRef](#)]
20. Panagopoulou, M.; Lambropoulou, M.; Balgkouranidou, I.; Nena, E.; Karaglani, M.; Nicolaidou, C.; Asimaki, A.; Konstantinidis, T.; Constantinidis, T.C.; Kolios, G.; et al. Gene promoter methylation and protein expression of BRMS1 in uterine cervix in relation to high-risk human papilloma virus infection and cancer. *Tumour Biol. J. Int. Soc. Oncodev. Biol. Med.* **2017**, *39*, 1010428317697557. [[CrossRef](#)]
21. Udomsinprasert, W.; Kitkumthorn, N.; Mutirangura, A.; Chongsrisawat, V.; Poovorawan, Y.; Honsawek, S. Association between Promoter Hypomethylation and Overexpression of Autotaxin with Outcome Parameters in Biliary Atresia. *PLoS ONE* **2017**, *12*, e0169306. [[CrossRef](#)]
22. Liu, X.; Peng, Y.; Wang, J. Integrative analysis of DNA methylation and gene expression profiles identified potential breast cancer-specific diagnostic markers. *Biosci. Rep.* **2020**, *40*. [[CrossRef](#)]
23. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)]
24. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210. [[CrossRef](#)]
25. Kim, Y.H.; Lee, H.C.; Kim, S.Y.; Yeom, Y.I.; Ryu, K.J.; Min, B.H.; Kim, D.H.; Son, H.J.; Rhee, P.L.; Kim, J.J.; et al. Epigenomic analysis of aberrantly methylated genes in colorectal cancer identifies genes commonly affected by epigenetic alterations. *Ann. Surg. Oncol.* **2011**, *18*, 2338–2347. [[CrossRef](#)]
26. Noreen, F.; Küng, T.; Tornillo, L.; Parker, H.; Silva, M.; Weis, S.; Marra, G.; Rad, R.; Truninger, K.; Schär, P. DNA methylation instability by BRAF-mediated TET silencing and lifestyle-exposure divides colon cancer pathways. *Clin. Epigenet.* **2019**, *11*, 196. [[CrossRef](#)]
27. Farkas, S.A.; Milutin-Gašperov, N.; Grce, M.; Nilsson, T.K. Genome-wide DNA methylation assay reveals novel candidate biomarker genes in cervical cancer. *Epigenetics* **2013**, *8*, 1213–1225. [[CrossRef](#)]
28. Wang, R.; Li, Y.; Du, P.; Zhang, X.; Li, X.; Cheng, G. Hypomethylation of the lncRNA SOX21-AS1 has clinical prognostic value in cervical cancer. *Life Sci.* **2019**, *233*, 116708. [[CrossRef](#)]
29. Conway, K.; Edmiston, S.N.; Parker, J.S.; Kuan, P.F.; Tsai, Y.H.; Groben, P.A.; Zedek, D.C.; Scott, G.A.; Parrish, E.A.; Hao, H.; et al. Identification of a Robust Methylation Classifier for Cutaneous Melanoma Diagnosis. *J. Investig. Dermatol.* **2019**, *139*, 1349–1361. [[CrossRef](#)]
30. Kirby, M.K.; Ramaker, R.C.; Roberts, B.S.; Lasseigne, B.N.; Gunther, D.S.; Burwell, T.C.; Davis, N.S.; Gulzar, Z.G.; Absher, D.M.; Cooper, S.J.; et al. Genome-wide DNA methylation measurements in prostate tissues uncovers novel prostate cancer diagnostic biomarkers and transcription factor binding patterns. *BMC Cancer* **2017**, *17*, 273. [[CrossRef](#)]
31. Najgebauer, H.; Liloglou, T.; Jithesh, P.V.; Giger, O.T.; Varro, A.; Sanderson, C.M. Integrated omics profiling reveals novel patterns of epigenetic programming in cancer-associated myofibroblasts. *Carcinogenesis* **2019**, *40*, 500–512. [[CrossRef](#)]
32. Shimada, S.; Mogushi, K.; Akiyama, Y.; Furuyama, T.; Watanabe, S.; Ogura, T.; Ogawa, K.; Ono, H.; Mitsunori, Y.; Ban, D.; et al. Comprehensive molecular and immunological characterization of hepatocellular carcinoma. *EBioMedicine* **2019**, *40*, 457–470. [[CrossRef](#)] [[PubMed](#)]
33. Cheishvili, D.; Stefanska, B.; Yi, C.; Li, C.C.; Yu, P.; Arakelian, A.; Tanvir, I.; Khan, H.A.; Rabbani, S.; Szyf, M. A common promoter hypomethylation signature in invasive breast, liver and prostate cancer cell lines reveals novel targets involved in cancer invasiveness. *Oncotarget* **2015**, *6*, 33253–33268. [[CrossRef](#)] [[PubMed](#)]
34. Lokk, K.; Modhukur, V.; Rajashekar, B.; Märtens, K.; Mägi, R.; Kolde, R.; Koltšina, M.; Nilsson, T.K.; Vilo, J.; Salumets, A.; et al. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol.* **2014**, *15*, r54. [[CrossRef](#)] [[PubMed](#)]
35. Braeuer, R.R.; Zigler, M.; Kamiya, T.; Dobroff, A.S.; Huang, L.; Choi, W.; McConkey, D.J.; Shoshan, E.; Mobley, A.K.; Song, R.; et al. Galectin-3 contributes to melanoma growth and metastasis via regulation of NFAT1 and autotaxin. *Cancer Res.* **2012**, *72*, 5757–5766. [[CrossRef](#)] [[PubMed](#)]
36. Yang, X.; Han, H.; De Carvalho, D.D.; Lay, F.D.; Jones, P.A.; Liang, G. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* **2014**, *26*, 577–590. [[CrossRef](#)]
37. Colaprico, A.; Silva, T.C.; Olsen, C.; Garofano, L.; Cava, C.; Garolini, D.; Sabedot, T.S.; Malta, T.M.; Pagnotta, S.M.; Castiglioni, I.; et al. TCGAAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **2016**, *44*, e71. [[CrossRef](#)]
38. Brenet, F.; Moh, M.; Funk, P.; Feierstein, E.; Viale, A.J.; Socci, N.D.; Scandura, J.M. DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS ONE* **2011**, *6*, e14524. [[CrossRef](#)]

39. Chandrashekar, D.S.; Basher, B.; Balasubramanya, S.A.H.; Creighton, C.J.; Ponce-Rodriguez, I.; Chakravarthi, B.; Varambally, S. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* **2017**, *19*, 649–658. [[CrossRef](#)]
40. Ando, M.; Saito, Y.; Xu, G.; Bui, N.Q.; Medetgul-Ernar, K.; Pu, M.; Fisch, K.; Ren, S.; Sakai, A.; Fukusumi, T.; et al. Chromatin dysregulation and DNA methylation at transcription start sites associated with transcriptional repression in cancers. *Nat. Commun.* **2019**, *10*, 2188. [[CrossRef](#)]
41. Deaton, A.M.; Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **2011**, *25*, 1010–1022. [[CrossRef](#)]
42. Perrakis, A.; Moolenaar, W.H. Autotaxin: Structure-function and signaling. *J. Lipid Res.* **2014**, *55*, 1010–1018. [[CrossRef](#)]
43. Fuss, B.; Baba, H.; Phan, T.; Tuohy, V.K.; Macklin, W.B. Phosphodiesterase I, a novel adhesion molecule and/or cytokine involved in oligodendrocyte function. *J. Neurosci.* **1997**, *17*, 9095–9103. [[CrossRef](#)]
44. Kanda, H.; Newton, R.; Klein, R.; Morita, Y.; Gunn, M.D.; Rosen, S.D. Autotaxin, an ectoenzyme that produces lysophosphatidic acid, promotes the entry of lymphocytes into secondary lymphoid organs. *Nat. Immunol.* **2008**, *9*, 415–423. [[CrossRef](#)]
45. Nema, R.; Shrivastava, A.; Kumar, A. Prognostic role of lipid phosphate phosphatases in non-smoker, lung adenocarcinoma patients. *Comput. Biol. Med.* **2021**, *129*, 104141. [[CrossRef](#)]
46. Wang, S.C.; Liao, L.M.; Ansar, M.; Lin, S.Y.; Hsu, W.W.; Su, C.M.; Chung, Y.M.; Liu, C.C.; Hung, C.S.; Lin, R.K. Automatic Detection of the Circulating Cell-Free Methylated DNA Pattern of GCM2, ITPRIPL1 and CCDC181 for Detection of Early Breast Cancer and Surgical Treatment Response. *Cancers* **2021**, *13*, 1375. [[CrossRef](#)]
47. Nauh, M.A.; Wu, X.X.; Okazoe, H.; Tsunemori, H.; Haba, R.; Abou-Zeid, A.M.; Saleem, M.D.; Inui, M.; Sugimoto, M.; Aoki, J.; et al. Expression of autotaxin and acylglycerol kinase in prostate cancer: Association with cancer development and progression. *Cancer Sci.* **2009**, *100*, 1631–1638. [[CrossRef](#)]
48. Magkrioti, C.; Oikonomou, N.; Kaffe, E.; Mouratis, M.A.; Xylourgidis, N.; Barbayianni, I.; Megadoukas, P.; Harokopos, V.; Valavanis, C.; Chun, J.; et al. The Autotaxin-Lysophosphatidic Acid Axis Promotes Lung Carcinogenesis. *Cancer Res.* **2018**, *78*, 3634–3644. [[CrossRef](#)]
49. Wu, J.M.; Xu, Y.; Skill, N.J.; Sheng, H.; Zhao, Z.; Yu, M.; Saxena, R.; Maluccio, M.A. Autotaxin expression and its connection with the TNF-alpha-NF-kappaB axis in human hepatocellular carcinoma. *Mol. Cancer* **2010**, *9*, 71. [[CrossRef](#)]
50. Maier, T.; Güell, M.; Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **2009**, *583*, 3966–3973. [[CrossRef](#)]
51. Chen, G.; Gharib, T.G.; Huang, C.C.; Taylor, J.M.; Misek, D.E.; Kardina, S.L.; Giordano, T.J.; Iannettoni, M.D.; Orringer, M.B.; Hanash, S.M.; et al. Discordant protein and mRNA expression in lung adenocarcinomas. *Mol. Cell. Proteom. MCP* **2002**, *1*, 304–313. [[CrossRef](#)] [[PubMed](#)]
52. Lemée, J.M.; Clavreul, A.; Aubry, M.; Com, E.; de Tayrac, M.; Mosser, J.; Menei, P. Integration of transcriptome and proteome profiles in glioblastoma: Looking for the missing link. *BMC Mol. Biol.* **2018**, *19*, 13. [[CrossRef](#)]
53. Arcos-Montoya, D.; Wegman-Ostrosky, T.; Mejía-Pérez, S.; De la Fuente-Granada, M.; Camacho-Arroyo, I.; García-Carrancá, A.; Velasco-Velázquez, M.A.; Manjarrez-Marmolejo, J.; González-Arenas, A. Progesterone Receptor Together with PKC α Expression as Prognostic Factors for Astrocytomas Malignancy. *OncoTargets Ther.* **2021**, *14*, 3757–3768. [[CrossRef](#)] [[PubMed](#)]
54. Tate, P.H.; Bird, A.P. Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr. Opin. Genet. Dev.* **1993**, *3*, 226–231. [[CrossRef](#)]
55. Farina, A.R.; Cappabianca, L.; Ruggeri, P.; Di Ianni, N.; Ragone, M.; Merolle, S.; Sano, K.; Stracke, M.L.; Horowitz, J.M.; Gulino, A.; et al. Constitutive autotaxin transcription by Nmyc-amplified and non-amplified neuroblastoma cells is regulated by a novel AP-1 and SP-mediated mechanism and abrogated by curcumin. *FEBS Lett.* **2012**, *586*, 3681–3691. [[CrossRef](#)] [[PubMed](#)]
56. Zhang, X.; Li, M.; Yin, N.; Zhang, J. The Expression Regulation and Biological Function of Autotaxin. *Cells* **2021**, *10*, 939. [[CrossRef](#)]
57. Héberlé, É.; Bardet, A.F. Sensitivity of transcription factors to DNA methylation. *Essays Biochem.* **2019**, *63*, 727–741. [[CrossRef](#)] [[PubMed](#)]
58. Chen, M.; O'Connor, K.L. Integrin alpha6beta4 promotes expression of autotaxin/ENPP2 autocrine motility factor in breast carcinoma cells. *Oncogene* **2005**, *24*, 5125–5130. [[CrossRef](#)] [[PubMed](#)]
59. Sioletic, S.; Czaplinski, J.; Hu, L.; Fletcher, J.A.; Fletcher, C.D.; Wagner, A.J.; Loda, M.; Demetri, G.D.; Sicinska, E.T.; Snyder, E.L. c-Jun promotes cell migration and drives expression of the motility factor ENPP2 in soft tissue sarcomas. *J. Pathol.* **2014**, *234*, 190–202. [[CrossRef](#)]
60. Benesch, M.G.; Tang, X.; Venkatraman, G.; Bekele, R.T.; Brindley, D.N. Recent advances in targeting the autotaxin-lysophosphatidate-lipid phosphate phosphatase axis in vivo. *J. Biomed. Res.* **2016**, *30*, 272–284. [[CrossRef](#)]
61. Barbayianni, E.; Kaffe, E.; Aidinis, V.; Kokotos, G. Autotaxin, a secreted lysophospholipase D, as a promising therapeutic target in chronic inflammation and cancer. *Prog. Lipid Res.* **2015**, *58*, 76–96. [[CrossRef](#)] [[PubMed](#)]

***ENPP2* Promoter Methylation Correlates with Decreased Gene Expression in Breast Cancer: Implementation as a Liquid Biopsy Biomarker**

Μεθυλίωση του υποκινητή του *ENPP2* συσχετίζεται με μειωμένη γονιδιακή έκφραση σε καρκινικό ιστό: εφαρμογή ως βιοδείκτης υγρής βιοψίας

Το γονίδιο *ENPP2* κωδικοποιεί για ένα σημαντικό ένζυμο της σύνθεσης του λυσοφωσφατιδικού οξέος (LPA). Προσφάτως περιγράψαμε τα προφίλ μεθυλίωσης του *ENPP2* τόσο σε υγιείς συνθήκες, όσο και σε πολλαπλές κακοήθειες και επιδείξαμε την συσχέτισή της με την αποκλίνουσα έκφραση του γονιδίου. Εδώ εστιάζουμε στον καρκίνο του μαστού (BrCa) αναλύοντας *in silico* δημοσίως διαθέσιμα σύνολα μεθυλίωσης, ώστε να εντοπίσουμε διαφορικά μεθυλιωμένα CpGs (DMCs) και να τα συσχετίσουμε με δεδομένα έκφρασης. Πολλαπλά CGs στο σώμα και στον υποκινητή του *ENPP2* εντοπίστηκαν διαφορικά μεθυλιωμένα μεταξύ BrCa και υγιών δειγμάτων μαστικού ιστού. Τα CGs του υποκινητή βρέθηκαν με περισσότερες μεθυλομάδες σε: δείγματα καρκίνου συγκρινόμενα με υγιή, δείγματα μεταστατικού καρκίνου σε σχέση με πρωτοπαθή και σε καρκινικά δείγματα σταδίου I σε σχέση με δείγματα ελέγχου. Αυτή η απορρυθμισμένη μεθυλίωση συσχετίστηκε με μειωμένη μεταγραφή του *ENPP2*. Επίσης, DMC του πρώτου εξωνίου ερευνήθηκε στο ελεύθερο DNA (cell free DNA) της κυκλοφορίας καρκινοπαθών και δειγμάτων ελέγχου. Η αυξημένη μεθυλίωση στα καρκινικά δείγματα επιβεβαίωσε τα υπολογιστικά δεδομένα. Επιπροσθέτως, διέφερε μεταξύ ομάδων ασθενών και συσχετίστηκε με την ύπαρξη πολλαπλών μεταστατικών θέσεων. Τα δεδομένα μας υποδεικνύουν ότι μεθυλίωση του υποκινητή του εν λόγω γονιδίου σταματά τη μεταγραφή του στον καρκίνο του μαστού, ενώ επίσης εισαγάγουν την μεθυλίωση του πρώτου εξωνίου ως έναν δυνητικό βιοδείκτη διάγνωσης και παρακολούθησης προσβάσιμο μέσω υγρής βιοψίας.



Article

ENPP2 Promoter Methylation Correlates with Decreased Gene Expression in Breast Cancer: Implementation as a Liquid Biopsy Biomarker

Maria Panagopoulou ^{1,2}, Andrianna Drosouni ¹, Dionysios Fanidis ³, Makrina Karaglani ^{1,2},
Ioanna Balgkouranidou ^{1,4}, Nikolaos Xenidis ⁴, Vassilis Aidinis ³ and Ekaterini Chatzaki ^{1,2,*}

¹ Laboratory of Pharmacology, Medical School, Democritus University of Thrace, 68100 Alexandroupolis, Greece; mpanagop@med.duth.gr (M.P.); androsouni@gmail.com (A.D.); mkaragla@med.duth.gr (M.K.); ioannabio@yahoo.it (I.B.)

² Institute of Agri-Food and Life Sciences, Hellenic Mediterranean University Research Center, 71410 Heraklion, Greece

³ Institute of BioInnovation, Biomedical Sciences Research Center Alexander Fleming, 16672 Athens, Greece; fanidis@fleming.gr (D.F.); aidinis@fleming.gr (V.A.)

⁴ Department of Medical Oncology, Democritus University of Thrace, 68100 Alexandroupolis, Greece; nxenidis@freemail.gr

* Correspondence: achatzak@med.duth.gr; Tel./Fax: +30-25-5103-0533

Abstract: Autotaxin (ATX), encoded by the ctonucleotide pyrophosphatase/phosphodiesterase 2 (*ENPP2*) gene, is a key enzyme in lysophosphatidic acid (LPA) synthesis. We have recently described *ENPP2* methylation profiles in health and multiple malignancies and demonstrated correlation to its aberrant expression. Here we focus on breast cancer (BrCa), analyzing in silico publicly available BrCa methylome datasets, to identify differentially methylated CpGs (DMCs) and correlate them with expression. Numerous DMCs were identified between BrCa and healthy breast tissues in the gene body and promoter-associated regions (PA). PA DMCs were upregulated in BrCa tissues in relation to normal, in metastatic BrCa in relation to primary, and in stage I BrCa in relation to normal, and this was correlated to decreased mRNA expression. The first exon DMC was also investigated in circulating cell free DNA (ccfDNA) isolated by BrCa patients; methylation was increased in BrCa in relation to ccfDNA from healthy individuals, confirming in silico results. It also differed between patient groups and was correlated to the presence of multiple metastatic sites. Our data indicate that promoter methylation of *ENPP2* arrests its transcription in BrCa and introduce first exon methylation as a putative biomarker for diagnosis and monitoring which can be assessed in liquid biopsy.

Keywords: autotaxin; *ENPP2*; methylation; breast cancer; liquid biopsy; expression; regulation



Citation: Panagopoulou, M.; Drosouni, A.; Fanidis, D.; Karaglani, M.; Balgkouranidou, I.; Xenidis, N.; Aidinis, V.; Chatzaki, E. *ENPP2* Promoter Methylation Correlates with Decreased Gene Expression in Breast Cancer: Implementation as a Liquid Biopsy Biomarker. *Int. J. Mol. Sci.* **2022**, *23*, 3717. <https://doi.org/10.3390/ijms23073717>

Academic Editor: Elena Bonanno

Received: 22 February 2022

Accepted: 25 March 2022

Published: 28 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Breast cancer (BrCa) is one of the most common cancers in the world among women [1]. Currently, early detection and new treatment options have improved the survival rate; however, clinical challenges still persist due to drug resistance and relapse being the leading cause of morbidity and mortality [2–4]. There is still an emerging need to define the biological mechanisms associated with the pathogenesis of BrCa and identify biomarkers and targets to improve treatment strategies.

The ATX-LPA signaling axis attracts growing interest in cancer research [5]. ATX is a secreted catalytically active glycoprotein that belongs to the ectonucleotide pyrophosphatase/phosphodiesterase (ENPP) family and is encoded by the *ENPP2* gene [5,6]. ATX has a lysophospholipase D activity and is mainly responsible for catalyzing the hydrolysis of extracellular LPC into LPA [5,6]. LPA then acts through at least six G-coupled receptors (GPCRs), known as LPAR1–6, and can activate various signaling pathways in almost every mammalian cell type [7]. Breast carcinogenesis was first linked to ATX and LPA signaling

back in 1995, by observations that ATX promotes proliferation of breast and ovarian cancer cells [8]. Since then, several studies have associated aberrant expression of ATX and LPA signaling with BrCa pathogenesis and metastatic progression [5,9–12].

Recently, a few studies reported that the ATX-LPA axis is governed by epigenetic regulation of the gene encoding ATX, ectonucleotide pyrophosphatase/phosphodiesterase 2 (*ENPP2*). DNA methylation is a well-studied epigenetic mechanism that regulates expression [13]. The identification of abnormal methylation in tissue or liquid biopsy has been correlated to cancer initiation and progression [14–17]. By employing an in silico approach, we have recently described *ENPP2* methylation profiles in health and malignancy, showing that methylation is an active level of ATX expression regulation in cancer. Increased methylation of promoter and first exon cytosine-guanine dinucleotides (CGs) and respective decreased *ENPP2* mRNA expression were found in prostate and lung cancers and were correlated to poor prognostic parameters [18].

Here, we focus in BrCa, presenting an in silico methylation and expression analysis of *ENPP2* followed by an experimental investigation in liquid biopsy. In specific, we explored the methylation status of *ENPP2* in BrCa and correlated it to its expression by using publicly available high-throughput methylation datasets from the Illumina methylation 450 bead-chip array, found in The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO) databases. Retrieved data were allocated into groups according to four important clinical endpoints related to prognosis and diagnosis to conduct differential methylation and expression analysis. *ENPP2* methylation, expression at protein and mRNA levels and survival were also estimated using the UALCAN platform. Identified differentially methylated *ENPP2* CGs were further validated in patient ccfDNAs to evaluate their potential for clinical implementation in liquid biopsies for the diagnosis and prognosis in BrCa.

2. Results

2.1. In Silico Analysis of *ENPP2* Methylation and Expression in BrCa

2.1.1. Differential Methylation and Expression Analysis between BrCa and Normal Breast Tissue

Raw methylome data from 520 BrCa (primary and metastatic) and 185 normal breast tissues were analyzed for the 14 CGs of *ENPP2* that the Infinium Human Methylation 450 k platform contains by means of RnBeads. In total, ten DMCS (FDR < 5×10^{-2}) were detected among breast tissues from healthy individuals and BrCa patients (Table 1). In promoter-associated (PA) regions known to be strongly associated with regulation of expression by methylation, i.e., transcription start site (TSS) and first Exon [19,20], all DMCs (cg04452959, cg02709432, cg02156680, cg06998282, and cg02534163) presented increased methylation in BrCa in relation to normal breast tissue. In the gene body, two CGs (cg00320790, cg20048037) were hypomethylated in BrCa and three CGs (cg09444531, cg26078665, cg23725583) were hypermethylated.

In order to address if the observed aberrant methylation of *ENPP2* in BrCa is associated with alterations in gene expression, we examined *ENPP2* mRNA levels in the same TCGA samples (GEO samples excluded as no expression data were available). Comparisons were made between 302 BrCa and 76 normal breast tissues. Results showed downregulation of expression in BrCa in relation to normal tissues (FC: -5.15 , FDR: 3.96×10^{-66}) (Table 2), indicating that the increased methylation of *ENPP2* in promoter and first exon regions is correlated with lower gene expression in BrCa. Discrete samples distribution based on methylation and expression is depicted in Figure 1. In specific, Figure 1A shows the distribution of BrCa and healthy samples in reduced dimensional space. Pathology is the main source of variation in both expression and methylation values. Correlation of *ENPP2* mRNA expression per CG methylation revealed statistically significant correlations, showed in Figure 2A and Table 3. In BrCa tissues, important correlations emerged only for PA CGs, as for TSS CGs (cg06998282, cg14409958) and first exon CG (cg02534163), methylation showed a reverse correlation with expression. These results further confirm

that increased methylation at these gene regions is associated with decreased expression. Analysis between *ENPP2* methylation and expression in normal breast tissues showed a negative correlation for two gene body CGs, namely cg07236691 and cg2372583, and a positive correlation for the gene body cg09444531 (Table 3).

Table 1. *ENPP2* DMCs identified via in silico analysis of BrCa and normal breast tissues.

CG ID	M β Value Normal	M β Value BrCa	$\Delta\beta$ Value	Methylation in BrCa	Gene Region	FDR
Normal breast tissue vs. BrCa						
cg00320790	0.96	0.95	0.01	Down	Body	5.97×10^{-4}
cg20048037	0.92	0.87	0.05	Down	Body	1.13×10^{-12}
cg09444531	0.77	0.79	-0.02	Up	Body	5.16×10^{-3}
cg26078665	0.77	0.84	-0.07	Up	Body	7.32×10^{-14}
cg23725583	0.85	0.92	-0.06	Up	Body	1.03×10^{-15}
cg02534163	0.06	0.53	-0.47	Up	1st Exon	3.15×10^{-91}
cg04452959	0.03	0.44	-0.41	Up	TSS200	4.56×10^{-80}
cg02709432	0.09	0.57	-0.48	Up	TSS200	6.71×10^{-73}
cg02156680	0.04	0.44	-0.39	Up	TSS1500	9.18×10^{-72}
cg06998282	0.09	0.62	-0.53	Up	TSS1500	9.54×10^{-76}
Primary vs. Metastatic BrCa						
cg20048037	0.87	0.82	0.06	Down	Body	3.99×10^{-2}
cg09444531	0.78	0.71	0.06	Down	Body	3.77×10^{-2}
cg26078665	0.86	0.79	0.07	Down	Body	9.09×10^{-4}
cg23725583	0.92	0.88	0.04	Down	Body	2.56×10^{-2}
cg02534163	0.55	0.74	-0.19	Up	1st Exon	1.26×10^{-4}
cg06998282	0.64	0.79	-0.15	Up	TSS1500	2.28×10^{-3}
Normal breast vs. stage I BrCa						
cg20048037	0.92	0.89	0.03	Down	Body	4.35×10^{-4}
cg09444531	0.77	0.80	-0.03	Up	Body	9.27×10^{-3}
cg26078665	0.78	0.86	-0.08	Up	Body	1.07×10^{-7}
cg23725583	0.86	0.93	-0.07	Up	Body	1.74×10^{-8}
cg02534163	0.06	0.55	-0.48	Up	1st Exon	2.45×10^{-49}
cg04452959	0.04	0.47	-0.43	Up	TSS200	5.53×10^{-38}
cg02709432	0.10	0.61	-0.51	Up	TSS200	1.14×10^{-38}
cg02156680	0.05	0.47	-0.43	Up	TSS1500	5.59×10^{-39}
cg06998282	0.10	0.66	-0.56	Up	TSS1500	2.93×10^{-35}
Early vs. Advanced BrCa						
cg01243251	0.94	0.92	0.014	Down	Body	3.10×10^{-2}
cg20162626	0.75	0.69	0.051	Down	Body	3.12×10^{-3}

Abbreviations: BrCa: breast cancer, DMCs: differentially methylated CpGs, FDR: false discovery rate, M β Value: mean β value, $\Delta\beta$ Value: difference between mean values.

Table 2. *ENPP2* differential expression analysis results based on TCGA data. $|FC| \geq 1.2$ and $FDR < 0.05$ are considered as thresholds of significant deregulation.

Compared Breast Groups	Fold Change	p-Value	FDR
Cancer_vs_Normal	-5.15	1.18×10^{-67}	3.96×10^{-66}
StageI_vs_Normal	-5.46	6.28×10^{-54}	3.43×10^{-52}
Advanced_vs_Early	1.20	1.23×10^{-2}	9.41×10^{-2}

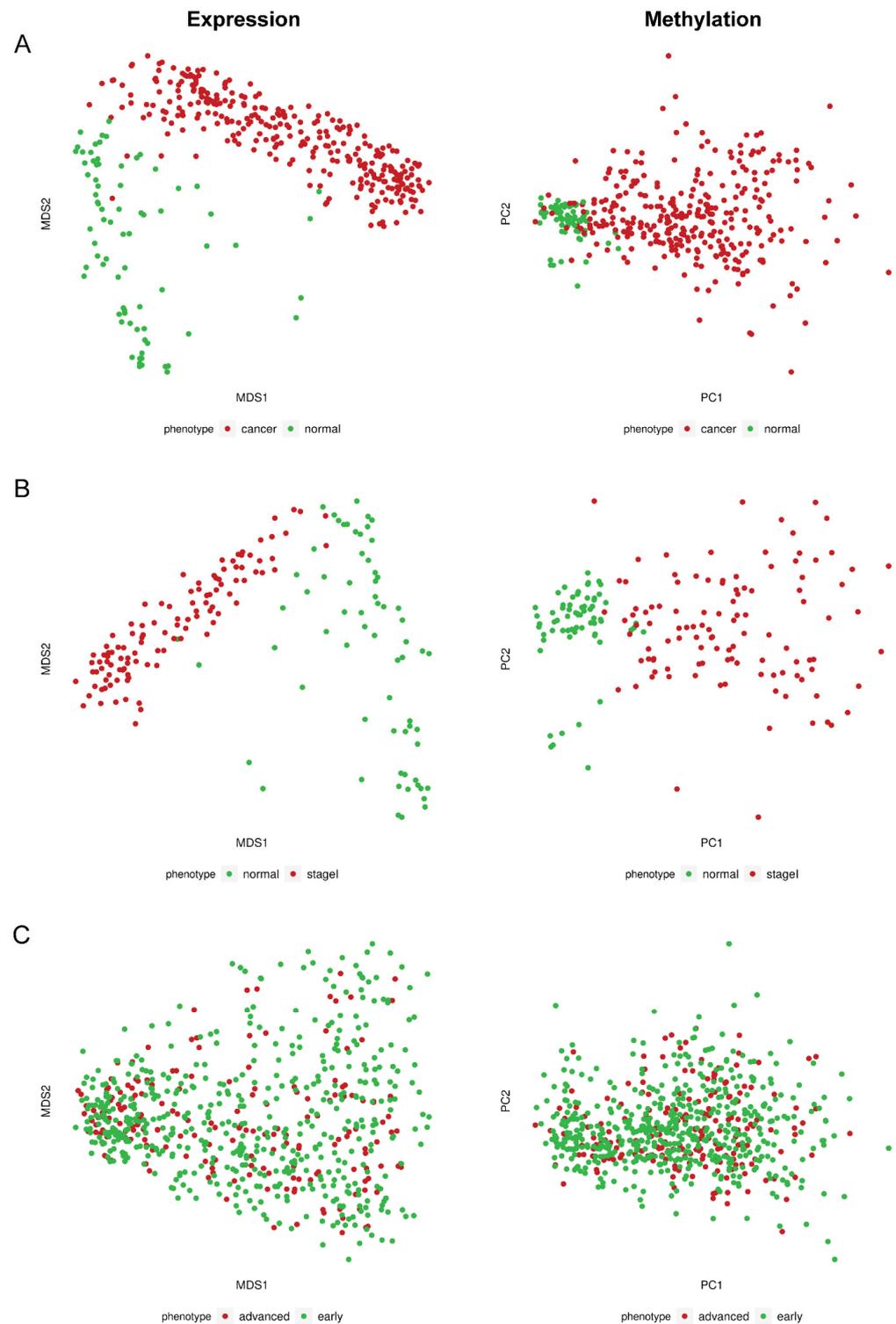


Figure 1. Dimensionality reduction plots for TCGA expression and methylation data. **(A)** MDS plot for normalized expression values of BrCa and normal breast tissue samples (**left**); PCA plot for level 3 beta methylation values of the same phenotypes (**right**). **(B)** MDS plot for normalized expression values of stage I BrCA and normal tissues (**left**); PCA plot for level 3 beta methylation values of the same phenotypes (**right**). **(C)** MDS plot for normalized expression values of advanced and early BrCA tissues (**left**); PCA plot for level 3 beta methylation values of the same phenotypes (**right**). MDS: Multidimensional scaling; PCA: principal component analysis.

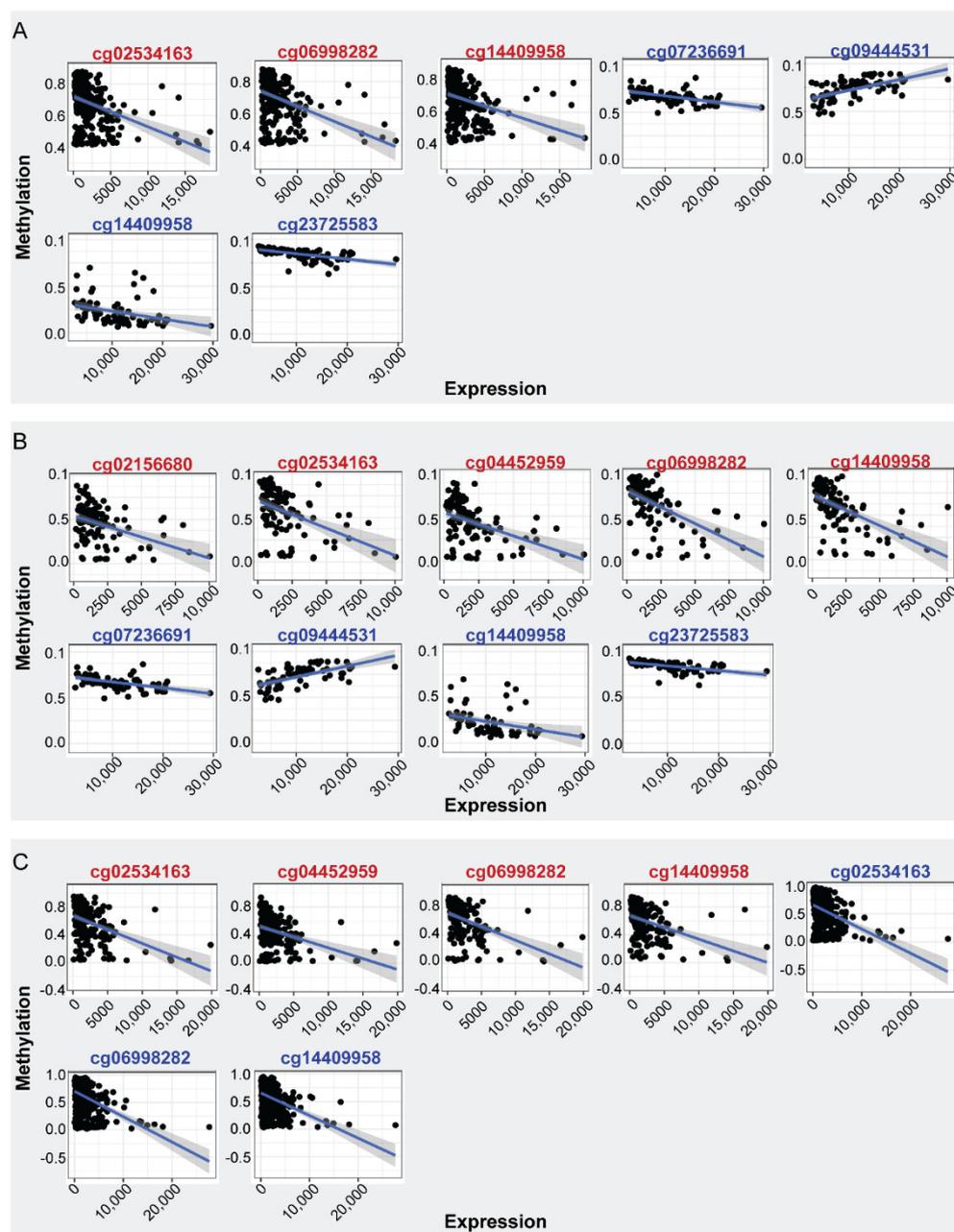


Figure 2. Spearman correlation of *ENPP2* CGs methylation and mRNA expression. CGs showing significant correlations are depicted ($|\rho| \geq 0.40$, $FDR < 5 \times 10^{-2}$). (A) Expression-methylation scatter plots of CG sites of BrCa (red) and normal (blue) samples, (B) expression-methylation scatter plots of CG sites of BrCa stage I (red) and normal (blue) samples, (C) expression-methylation scatter plots of CG sites of advanced (red) and early (blue) BrCa samples.

2.1.2. Differential Methylation Analysis between Primary and Metastatic BrCa

Methylomes of primary BrCa were analyzed in comparison to those from metastatic BrCa in order to detect changes in *ENPP2* related to metastatic transformation. Raw data from 132 primary cancers and 31 cancers with distant metastasis were analyzed using RnBeads and 6 DMCs out of a total of 14 CGs ($FDR < 5 \times 10^{-2}$) were detected (Table 1). Four of them were located at the gene body (cg20048037, cg09444531, cg26078665, cg23725583) presenting lower methylation and two were located in the TSS and first exon (cg06998282 and cg02534163, respectively) showing upregulation in metastatic in relation

to primary BrCa. These observations suggest an involvement of *ENPP2* methylation in BrCa progression and metastasis.

Table 3. Spearman correlation coefficient between *ENPP2* CG methylation and mRNA expression for each of our clinical endpoints. CGs showing significant correlations are depicted ($|\rho| \geq 0.40$, $FDR < 5 \times 10^{-2}$).

Tissue	CG	BrCa vs. Normal		FDR	Correlation
		Gene Region	Rho		
BrCa	cg02534163	First Exon	−0.40	1.18×10^{-12}	Negative
	cg06998282	TSS1500	−0.42	1.53×10^{-13}	Negative
	cg14409958	TSS1500	−0.42	2.10×10^{-13}	Negative
Normal	cg09444531	Body	0.62	4.77×10^{-06}	Positive
	cg23725583	Body	−0.70	3.02×10^{-11}	Negative
	cg07236691	Body	−0.55	1.27×10^{-08}	Negative
	cg14409958	TSS1500	−0.52	9.53×10^{-07}	Negative
Stage I BrCa vs. Normal					
Stage I	cg02534163	First Exon	−0.46	1.54×10^{-06}	Negative
	cg04452959	TSS200	−0.44	3.74×10^{-06}	Negative
	cg14409958	TSS1500	−0.64	3.29×10^{-13}	Negative
	cg02156680	TSS1500	−0.42	1.28×10^{-05}	Negative
	cg06998282	TSS1500	−0.63	6.04×10^{-13}	Negative
Normal	cg09444531	Body	0.64	3.25×10^{-08}	Positive
	cg23725583	Body	−0.66	1.36×10^{-08}	Negative
	cg07236691	Body	−0.53	1.40×10^{-05}	Negative
	cg14409958	TSS1500	−0.51	4.25×10^{-05}	Negative
Early vs. Advanced BrCa					
Early	cg02534163	First Exon	−0.42	2.56×10^{-23}	Negative
	cg06998282	TSS1500	−0.47	4.588×10^{-29}	Negative
	cg14409958	TSS1500	−0.46	5.37×10^{-28}	Negative
Advanced	cg02534163	First Exon	−0.43	1.23×10^{-09}	Negative
	cg04452959	TSS200	−0.41	5.47×10^{-09}	Negative
	cg14409958	TSS1500	−0.48	4.72×10^{-12}	Negative
	cg06998282	TSS1500	−0.49	4.18×10^{-12}	Negative

Abbreviations: FDR: false discovery rate, TSS: Transcription Start Site.

2.1.3. Differential Methylation and Expression Analysis between Stage I BrCa and Normal

In order to address if aberrant *ENPP2* methylation is an early effect in the breast carcinogenic process, methylome raw data from 136 stage-I BrCa and 111 normal breast tissues were subjected to RnBeads differential methylation analysis. A total of 9 out of the 14 studied CGs were DMCs between stage I BrCa and normal tissues (Table 1). All but one gene body DMCs and all five PA DMCs showed increased methylation in stage I BrCa in relation to normal tissues.

Differential mRNA expression analysis between 111 Stage I BrCa and 66 Normal breast tissue samples showed downregulation of expression in Stage I cancer (FC:−5.46, FDR: 3.43×10^{-52}) (Table 2), similarly with findings of the analysis between all BrCa samples and normal tissues. Dimensionality reduction plot (Figure 1B) depicts excellent separation of Stage I BrCa and normal samples based on level 3 methylation and normalized expression values. Interestingly, a negative correlation between methylation and expression was noted for all TSS (cg02156680, cg04452959, cg06998282, cg14409958) and first exon CGs (cg02534163) in BrCa samples and no correlation emerged for gene body CGs (Table 3). A different pattern was observed for control samples as methylation of three gene body CGs, namely cg07236691 (negatively), cg23725583 (negatively), and cg09444531 (positively), were

found to be correlated to mRNA expression. Finally, one TSS CG (cg14409958) presented reverse correlation between methylation and expression (Table 3).

2.1.4. Differential Methylation and Expression Analysis between Early- and Advanced-Stage BrCa

In order to detect important methylation events related to the progression of BrCa to advanced-stage disease, we conducted an analysis of raw methylome data from 521 early (stage I, II) and 221 advanced (stage III) BrCa patients. Only two DMCs (cg01243251, cg20162626) were identified in the gene body region, showing a slight but still statistically significant decrease in methylation in advanced in relation to early BrCa. No difference was observed in TSS or first exon CGs (Table 1). In accordance to methylation, no difference in mRNA expression was observed between 519 stage I and II cancers and 191 stage III cancers (FC: 1.20, FDR: 9.41×10^{-2}) (Table 2). Last, in dimensionality reduction plots (Figure 1C) there was no separation of early and advanced stage BrCa samples based on detected features' methylation and expression.

As expected, correlation analysis revealed important relationships for PA CGs. In specific, in advanced cancer, methylation of three TSS (cg04452959, cg06998282, cg14409958) and one first exon CG (cg02534163) was negatively correlated with *ENPP2* expression (Figure 2C and Table 3). Same tendency was observed in the early cancers except for the case of cg04452959 (Figure 2C and Table 3).

Cumulatively, previous analysis indicated that *ENPP2* methylation is associated with the malignant transformation of breast cells and metastasis. However, minimal methylation changes were noted between stage I/II and stage III BrCa. Similarly, downregulation of *ENPP2* expression was noted in BrCa samples in relation to control but not between early and advanced stage.

2.1.5. Differential Methylation Analysis between BrCa Cancer Types

We also examined differences in the methylation of *ENPP2* between BrCa cancer types. In particular, differential methylation analysis was performed between 473 invasive ductal and 186 invasive lobular BrCa (accounting for 90.2% of available TCGA cases). Three CGs of *ENPP2* were found differentially methylated (Table 4), showing small but statistically significant differences. In specific, body CGs were either hyper-(cg01243251) or hypomethylated (cg20048037) in ductal in relation to lobular BrCa. A third CG located at the TSS1500 (cg02156680) presented downregulation of methylation in the ductal type.

Table 4. *ENPP2* DMCs identified between ductal and lobular BrCa via in silico analysis of TCGA cases.

CG ID	M β Value Ductal Cancer	M β Value Lobular Cancer	$\Delta\beta$ Value	Methylation in Ductal Cancer	Gene Region	FDR
cg01243251	0.94	0.93	0.01	Up	Body	1.53×10^{-2}
cg20048037	0.86	0.90	-0.04	Down	Body	1.11×10^{-2}
cg02156680	0.47	0.52	-0.04	Down	TSS1500	2.52×10^{-2}

Abbreviations: BrCa: breast cancer, DMCs: differentially methylated CpGs, FDR: false discovery rate, M β Value: mean β value, $\Delta\beta$ Value: difference between mean values.

2.1.6. In Silico Analysis of *ENPP2* Methylation in BrCa ccfDNA Data

The analysis of a ccfDNA dataset (GSE1222126) revealed eight *ENPP2* DMCs (Table 5) between BrCa patient-derived samples and healthy individuals. In specific, four (cg04452959, cg02156680, cg06998282, cg14409958) out of five studied TSS CGs and one first exon CG (cg02534163) were found hypermethylated in BrCa ccfDNAs in relation to control. In the gene body, two (cg07236691, cg20162626) and one (cg20048037) CGs were hypomethylated and hypermethylated, respectively, in BrCa. It has to be noted that four DMCs in PA regions were common between BrCa ccfDNA and breast tissue samples (clinical endpoint: BrCa vs. normal), suggesting that ccfDNA may reflect the methylation status of the tumor.

Table 5. *ENPP2* DMCs identified in ccfDNA of BrCa patients and healthy individuals via in silico analysis.

CG ID	M β Value BrCa	M β Value Normal	$\Delta\beta$ Value	Methylation in BrCa	Location	FDR
cg07236691	0.583	0.817	−0.234	Down	Body	3.04×10^{-3}
cg20048037	0.711	0.331	0.380	Up	Body	5.08×10^{-5}
cg20162626	0.489	0.814	−0.325	Down	Body	3.18×10^{-4}
cg02534163	0.802	0.206	0.596	Up	1st Exon	6.30×10^{-11}
cg04452959	0.620	0.020	0.599	Up	TS200	2.51×10^{-13}
cg02156680	0.515	0.028	0.487	Up	TSS1500	1.48×10^{-13}
cg06998282	0.772	0.057	0.715	Up	TSS1500	1.61×10^{-8}
cg14409958	0.748	0.053	0.695	Up	TSS1500	2.24×10^{-6}

Abbreviations: BrCa: breast cancer, DMCs: differentially methylated CpGs, FDR: false discovery rate, M β Value: mean β value, $\Delta\beta$ Value: difference between mean values.

2.2. *ENPP2* Methylation, Expression and Survival Analysis by UALCAN

In order to further verify our findings, we conducted *ENPP2* expression, methylation, and survival analysis in BrCa using the UALCAN platform. Analysis confirmed above results, as promoter methylation level of *ENPP2* was increased in primary tumor tissues of BrCa patients in relation to normal ($p < 1 \times 10^{-12}$) as depicted in Figure 3A. Next, expression analysis showed downregulation of *ENPP2* mRNA expression in primary tumor tissues in relation to normal tissues ($p = 1.6 \times 10^{-12}$) as depicted in Figure 3B. Similarly, protein expression analysis showed downregulation in primary tumors in relation to normal tissues (Figure 3E, $p = 4.5 \times 10^{-12}$). Methylation and expression results by UALCAN strengthen our findings, showing that the *ENPP2* gene is methylated in BrCa and this is related to lower expression, suggesting a causative relationship and a methylation regulatory mechanism in BrCa. Finally, survival analysis did not reveal any statistical significance as depicted in Supplementary Figure S1.

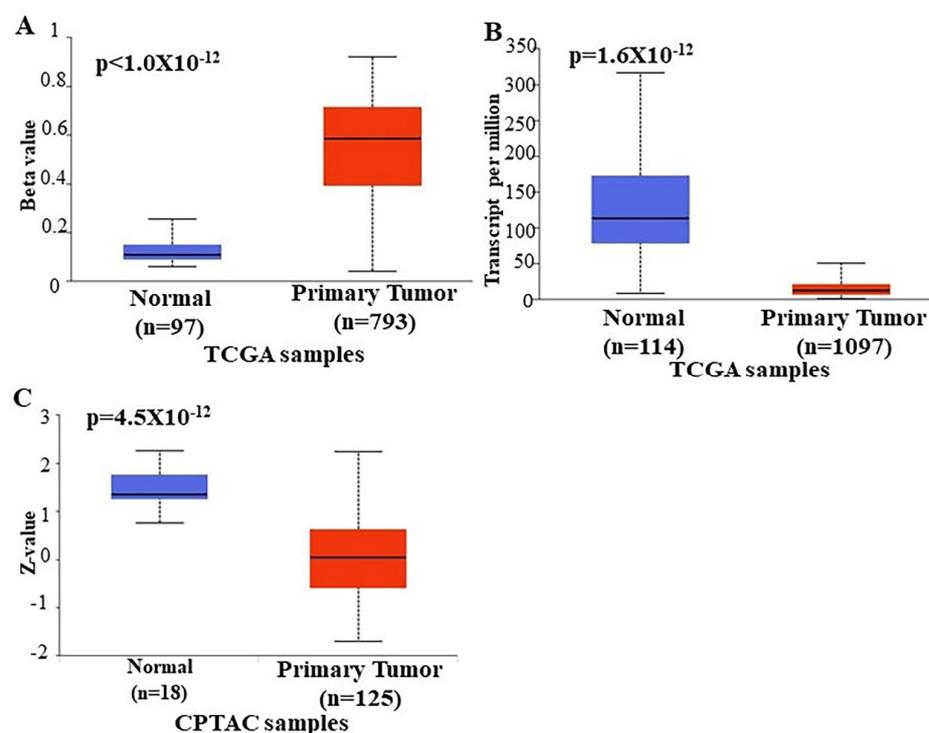


Figure 3. Analysis of *ENPP2* (A) promoter DNA methylation, (B) mRNA expression, and (C) ATX protein expression between primary BrCa tumors and normal tissues using the UALCAN platform. Abbreviations: TCGA: The Cancer Genome Atlas, CPTAC: Clinical Proteomic Tumor Analysis Consortium, BrCa: breast cancer.

2.3. Methylation Analysis of *ENPP2* in ccfDNA from BrCa Patients

Following the in silico analysis, the methylation of *ENPP2* was investigated in BrCa patient-derived ccfDNAs and was compared to their healthy counterparts, using qMSP, in order to test clinical applicability in liquid biopsy. Primers were designed to include the cg02534163 of the first exon of *ENPP2*, a CG identified as a DMC in the in silico analysis in both tissues, and ccfDNA and could be exploited as a biomarker in BrCa.

ENPP2 methylation was investigated in ccfDNAs isolated from 52 adjuvant, 19 metastatic, and 15 neoadjuvant BrCa patients and 20 healthy individuals (control). Methylation was detected more often in ccfDNA of BrCa patients than in healthy individuals in a statistically significant manner ($p = 1 \times 10^{-2}$) (Figure 4A). In specific, methylated *ENPP2* was detected in 21 out of 46 (45.6%) of control samples and in 62 out of 86 (72.1%) of BrCa samples. Between BrCa groups, methylation was detected in 71.1% (36/52), 73.6% (14/19), and 80% (12/15) of adjuvant, metastatic, and neoadjuvant groups, respectively, showing statistically significant differences between groups ($p = 2 \times 10^{-2}$). In specific, when methylation positives of each group were compared separately, statistically significant correlations emerged between control and adjuvant ($p = 9 \times 10^{-3}$), control and metastatic ($p = 3.6 \times 10^{-2}$), control and neoadjuvant ($p = 2 \times 10^{-2}$), but not between BrCa groups.

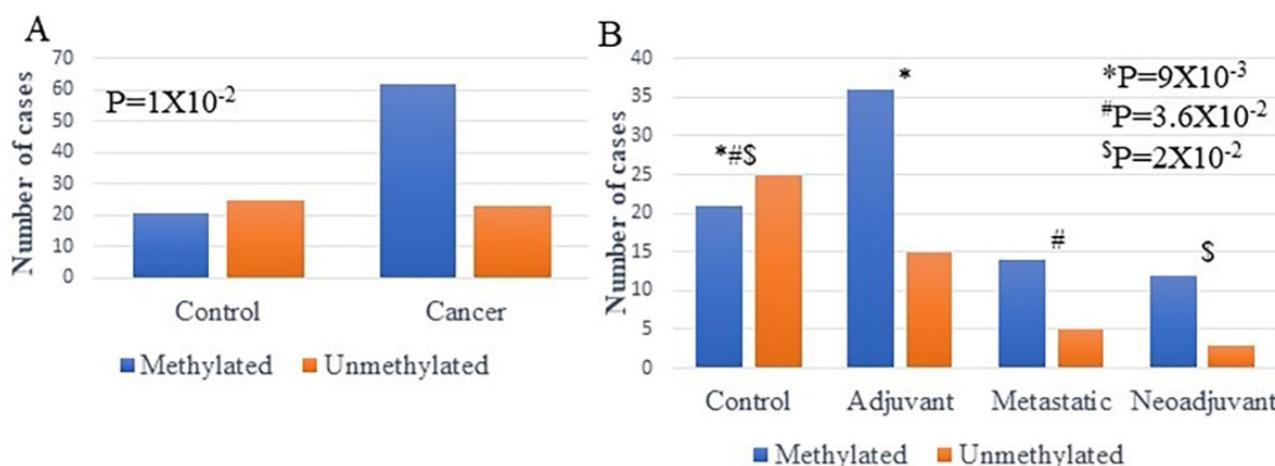


Figure 4. Methylation of *ENPP2* estimated by qMSP in ccfDNA (A) from BrCa patients and healthy individuals (B) from each group separately (adjuvant, metastatic, neoadjuvant, and healthy). Abbreviations: ccfDNA: circulating cell-free DNA; qMSP: quantitative methylation-specific PCR; CRC: colorectal cancer. * Control in relation to Adjuvant, # Control in relation to Metastatic, \$ Control in relation to Neoadjuvant.

Next, methylation levels were also measured. Methylation levels were found elevated in ccfDNA of BrCa patients compared to ccfDNA of healthy individuals, but no statistically significant correlation emerged ($p = 8 \times 10^{-2}$) (Figure 5A). Between groups, significantly increased levels of *ENPP2* methylation were found in the neoadjuvant group as compared to the control group ($p = 8 \times 10^{-4}$) and the adjuvant group ($p = 1 \times 10^{-3}$) (Figure 5B). This result could be due to the fact that neoadjuvant patients have increased tumor burden in relation to adjuvant patients, having their tumor removed. No other statistically significant difference in methylation levels were observed between the other studied groups (Control vs. Adjuvant: 4.3×10^{-1} , Control vs. Metastatic: 4.5×10^{-1} , Adjuvant vs. Metastatic: 8.4×10^{-1} , Metastatic vs. Neoadjuvant: 5.5×10^{-1}) (Figure 5B). A significant correlation was found in the metastatic group, as patients with one metastatic site presented lower methylation levels than those having more metastatic sites ($p < 1 \times 10^{-2}$) (Figure 5C). Finally, no other correlation emerged between *ENPP2* methylation and grade, nodal status, tumor size, or age.

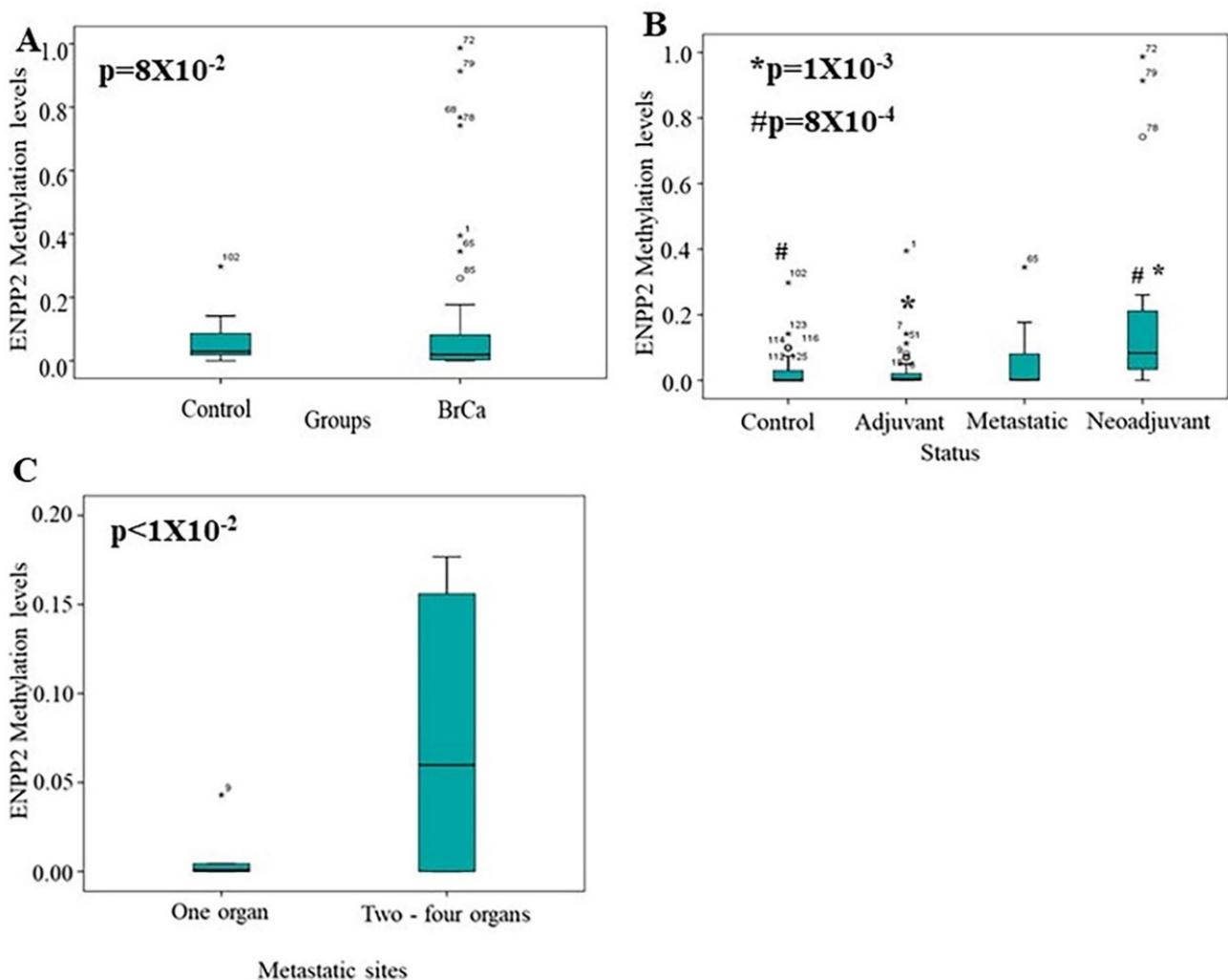


Figure 5. Median methylation levels of *ENPP2*. Boxplots depict methylation levels in ccfDNA of (A) BrCa compared to control group and (B) in each studied group separately; (C) BrCa patients with one metastatic site compared to those with two or more. Abbreviations: BrCa: breast cancer, * Adjuvant in relation to Neoadjuvant, # Control in relation to Neoadjuvant.

3. Discussion

ATX is a well-known enzyme responsible for generating lysophosphatidic acid (LPA) and dysregulation of its expression has been linked to several pathologies and to cancer [5,6]. ATX's encoding gene *ENPP2*, is epigenetically regulated, as aberrant methylation patterns were described in five different cancer types and were correlated to mRNA expression. Furthermore, increased methylation of *ENPP2* was connected to poor prognostic parameters [18]. In most cancer types, ATX is increased, but BrCa cells express little ATX. Still, ATX from its microenvironment plays an important role in BrCa development, promoting cell proliferation, migration, and survival, and is also regarded as a potential target for therapy or increased chemotherapeutic sensitivity [21].

In the present study, we focus on *ENPP2* methylation and expression in BrCa. We first adopted a bioinformatic approach using publicly available datasets of BrCa tissues and ccfDNA. Among 10 DMCs identified between BrCa and normal tissues, all TSS and 1st Exon DMCs presented increased methylation in BrCa. These gene regions are known to be strongly associated with regulation of expression by methylation [19,20] indicating epigenetic arrest of *ENPP2* transcription in BrCa. Indeed, expression analysis showed decreased transcription in BrCa in relation to normal tissues, and increased methylation of all PA CGs was reversely correlated to mRNA expression, results also confirmed by

analysis in the UALCAN platform. These results agree with our previous findings in HCC, melanoma, CRC, LC, and PC, showing *ENPP2* hypermethylation in PA and decreased expression [18].

In addition, *ENPP2* is one of 11 genes spanning the 8q12.1-q24.22 genomic region found to be differentially methylated and expressed in invasive breast carcinomas harboring the 8p11-p12 amplicon by integrative analysis. In accordance to our results, *ENPP2* was the only gene showing lower expression levels and hypermethylation despite amplification of the 8p11-p12 amplicon [22]. In fact, one of the DMCs identified here in BrCa and previously in HCC and PC, located in TSS1500 (cg02156680) was included in an eight-feature methylation-based breast-cancer specific signature constructed by integrative analysis of genome-wide DNA methylation and was shown to reliably separate BrCa from normal samples [23]. In contrast, the whole *ENPP2* gene was not included in the methylation-based biosignatures of translational relevance built via automated machine learning analysis of BrCa whole methylome datasets and was not identified as a top rated DMG [24].

When compared to healthy tissues, stage I BrCa showed hypermethylation in PA CGs which were correlated to downregulation of expression, indicating an early event in BrCa. However, when comparison was made between early (I, II) and advanced (III) stages of BrCa, no difference was observed in the methylation pattern of PA CGs, but only in two gene body CGs, and no changes in mRNA expression, suggesting minor *ENPP2* methylation events in the course of the disease. Another study based on TCGA datasets analysis reported *ENPP2* as one of 66 significantly hypermethylated genes with logFC > 1.8 between Stage I–III BrCa [25].

Comparison between primary and metastatic BrCa revealed six *ENPP2* DMCs. Among them, two were in TSS and first exon showing hypermethylation of *ENPP2* in metastatic BrCa, implying a participation in the metastatic cascade. Increased expressions of ATX in the stroma is associated with aggressiveness of human BrCa in women [26], whereas *ENPP2* is one of the 40–50 most up-regulated genes in metastatic solid tumors [5]. Unfortunately, expression data from metastatic samples were not available to allow correlation with methylation and deeper understanding in the process of metastasis. It has to be noted that in our previous work, *ENPP2* hypermethylation in lung cancer was correlated with advanced cancer stage [18].

Small differences in *ENPP2* methylation were observed between ductal and lobular BrCa pathological types, and no further analysis was possible due to lack of relevant clinical information in the available datasets and small representation from other BrCa types. GEO datasets, for example, do not include the cancer type as a parameter for each case. This underlines the significance of the integrity of information provided in the archived datasets to allow full exploitation of readings towards clinical relevance.

ENPP2 hypermethylation of PA associated CGs was also detected in methylome datasets of ccfDNAs from BrCa patients in relation to ccfDNA from healthy individuals, presenting a similar profile as in the case of tissue samples. Taken together, these findings suggest that assessing methylation in ccfDNA can dynamically reflect methylation events of the tumor. This notion is further supported by our recent in vitro study showing that the methylation profile of ccfDNA released by breast and cervical cancer cell lines is identical to their genomic DNA [27]. Similarly, in CRC, we detected identical methylation profiles of corticotropin releasing factor receptor genes in tumors as in patient ccfDNAs [16].

We further evaluated *ENPP2* methylation in ccfDNA of BrCa patients in order to examine its clinical value as a biomarker. The first exon cg02534163 was targeted in a qMSP assay, chosen because it was identified as a DMC in all clinical endpoints examined, except in the case of early vs. advanced disease, presenting the highest FDR. Analysis showed that *ENPP2* hypermethylation was detected more often in ccfDNA of BrCa patients than in healthy individuals in a statistically significant manner. In a previous study addressing *ENPP2* methylation in ccfDNA from 22 healthy and 45 Taiwanese BrCa patients, no significant differences were found [28], although *ENPP2* methylation showed a twofold increase in BrCa in relation to adjacent normal tissue. Methodological differences or even

population genetic variations might explain these different findings. Importantly, in our study, ccfDNA methylation levels of *ENPP2* were also elevated in the neoadjuvant and metastatic groups of patients in relation to adjuvant and control group of patients. This result could be due to the fact that in the new adjuvant and metastatic groups, patients still have a significant tumor burden. Our bioinformatic analysis also showed that *ENPP2* methylation is increased in metastasis in relation to primary cancers. In addition to that, according to our experimental analysis, patients having two or more metastatic foci presented more increased *ENPP2* methylation levels than those patients having a distant metastasis in one organ. Cumulatively, our experimental results are in accordance with those from bioinformatic analysis showing hypermethylation of *ENPP2* in BrCa tissue and ccfDNA and a correlation with cancer aggressiveness and metastasis, suggesting its potential as a novel circulating biomarker in BrCa.

A limitation of our study is the small number of patients enrolled in the experimental part not allowing significant correlations between *ENPP2* methylation levels and clinicopathological features such as grade and tumor size to emerge. Future validation in a larger group of patients should be conducted in order to confirm its clinical value. Furthermore, studies employing in vitro models should clarify the connection between methylation and expression during the carcinogenic process.

4. Materials and Methods

4.1. Bioinformatic Analysis of *ENPP2* in BrCa

4.1.1. Data Sources

Raw DNA methylation data from BrCa tissues and normal breast tissues as well as the corresponding clinical and demographic data were obtained from The Cancer Genome Atlas (TCGA) [29] and Gene Expression Omnibus (GEO) [30] databases. TCGA case inclusion criteria were: 1. Platform: Infinium Human Methylation 450 K bead-chip 2. Primary site: breast; 3. Project: TCGA-BRCA; 4. Gender: female; 5. Age at diagnosis: 26–80 years; 6. Race: white, black or African American, Asian, and not reported. A total of 730 cases were downloaded. As for the BrCa type at diagnosis, the majority of the cases were invasive ductal (64.7% of total cases) or lobular (25.5% of total cases) BrCa, 9.8% of cases were of eight other BrCa types (e.g., secretory, tubular, papillary, and others). The GEO database was searched using ‘Breast cancer’, ‘Metastatic Breast cancer’, ‘cell free DNA’ as keywords and ‘Methylation profiling by array’ as study type. In total, 96 studies were found. Those using the Infinium Human Methylation 450 K bead-chip array and providing adequate raw and clinical data were selected for further analysis, i.e., five studies, namely GSE72245, GSE72251 [31], GSE88883 [32], GSE108576 [33], GSE74214 and GSE122126 [34]. Analysis of *ENPP2* methylation in tissues was performed against 4 major clinically relevant endpoints, as presented in Table 6. Analysis of *ENPP2* methylation was also performed in a ccfDNA dataset (GSE122126) [34] including three BrCa ccfDNA samples and two from healthy individuals.

Table 6. Comparisons, endpoints, study group characteristics, and clinical significance of the tissue datasets used in the bioinformatic analysis. Abbreviations: BrCa = breast cancer, NR = Not Relevant.

Study Groups	Tissues	Age (Years) Median (Range)	Stage	Significance
1. BrCa vs. Normal	520 BrCa (primary and metastatic)	49 (26–80)	102 Stage I 264 Stage II 114 Stage III 40 Stage IV	Diagnosis
	185 Normal	47 (26–80)	NR	
2. Primary vs. Metastatic BrCa	132 Primary BrCa	55 (47–55)	22 Stage I 75 Stage II 35 Stage III	Diagnosis/Prognosis
	31 Metastatic BrCa	54 (41–80)	31 Stage IV	
3. Stage I BrCa vs. Normal	136 Stage I BrCa	54 (27–80)	136 Stage I	Diagnosis/Prognosis
	111 Normal	58 (29–80)	NR	
4. Early vs. Advanced BrCa	521 Early BrCa	58 (26–80)	115 Stage I 406 Stage II	Diagnosis/Prognosis
	221 Advanced BrCa	55 (27–80)	221 Stage III	

4.1.2. Data Preprocessing and DNA Methylation Analysis

Raw DNA methylation data (IDAT files) and sample annotation files were subjected to the Bioconductor R package RnBeads v2.0 [35]. RnBeads is a software tool suitable for large-scale analysis, interpretation, and visualization of DNA methylation data. In our workflow, *ENPP2* CGs were chosen as the genomic region of interest and were analyzed for each of the four endpoints, as previously reported by our team [16,24]. Beta methylation values are expressed as decimal values between 0.0 (no methylation) and 1.0 (full methylation). DMCs (DMCs) for *ENPP2* were identified based on the false discovery rate (FDR-adjusted p -value $< 5.00 \times 10^{-2}$).

4.1.3. Differential Expression Analysis and Expression—Methylation Correlation

Raw RNA-seq (Illumina HiSeq) paired with level 3 methylation data described above (Table 6) were obtained from the TCGA database. No expression data were available for the GEO retrieved methylation datasets. In detail, 66 Normal, 302 BrCa, 111 Stage I, 191 advanced, and 519 early-stage samples were obtained. Expression data were EDASeq normalized and quantile filtered post to differential expression analysis using the edgeR package. Absolute fold change ≥ 1.2 and adjusted p -value $\leq 5 \times 10^{-2}$ were selected as thresholds of significant differential expression. Multi-dimensional scaling (MDS) plots of normalized expression and principal component analysis (PCA) of level 3 methylation data were created to visualize sample separation according to the phenotype attributed. Normalized expression and level 3 methylation data were correlated using the Spearman method. Absolute rho value ≥ 0.4 and adjusted p value $\leq 5 \times 10^{-2}$ were set as significant correlation thresholds. All the above manipulations were performed with TCGA biolinks package version 2.18.0 [36] and R version 4.0.4.

4.1.4. Expression, Methylation and Survival Analysis Using the UALCAN Platform

In order to further verify our results, we used the UALCAN platform [37] that enable researchers to analyze cancer archived omics data. We performed expression, promoter methylation, and survival analysis of *ENPP2* gene in BrCa and corresponding controls. According to UALCAN, different beta value cut-offs have been considered to indicate hypermethylation (Beta value: 0.7–0.5) or hypo-methylation (Beta-value: 0.3–0.25). For mRNA expression, methylation, and survival, we used TCGA gene analysis, and the screening conditions were as follows: gene “*ENPP2*”, TCGA dataset “Breast Cancer” and then we used “expression”, “methylation”, and “survival” as links for analysis. Protein expression analysis was performed using the Clinical Proteomic Tumor Analysis Consortium (CPTAC)

datasets, Z-values represent standard deviations from the median across samples for the given cancer type. Log2 Spectral count ratio values from CPTAC were first normalized within each sample profile then normalized across samples.

4.2. Methylation Analysis of ENPP2 in BrCa Liquid Biopsies

4.2.1. Study Groups and Clinical Samples

The study was approved by the Scientific Board of the University General Hospital of Evros (PGNE), following assessment by Ethics Committee (decision 663/08.08.16), and was conducted according to the ethical principles of the 1964 Declaration of Helsinki and its later amendments. All patients participated after signing a voluntary informed consent.

Blood samples were collected from 86 BrCa patients who visited the Department of Medical Oncology of PGNE and were allocated to three groups: (a) 52 patients having recently (within the previous month) undergone surgery for primary BrCa, exactly before the initiation of adjuvant therapy (adjuvant group), (b) 15 patients upon diagnosis for BrCa, having no previous surgery, before the initiation of neo-adjuvant therapy (neo-adjuvant group), (c) 19 patients upon diagnosis for metastatic disease before the initiation of first-line chemotherapy (a combination of Taxane/Anthracyclines) (metastatic group). Pathological BrCa type was invasive ductal carcinoma for all patients enrolled in the study. The available clinicopathological features for all patient groups are presented in Table 7. Five-year follow-up data were also available. In the adjuvant group, 10 (19.23%) patients have died as a consequence of their disease progression and respective numbers in the metastatic and neo-adjuvant BrCa groups were 11 (57.89%) and 5 (33.33%), respectively.

Table 7. Demographic and clinicopathological characteristics of BrCa and control groups.

Group	Total	Adjuvant	Metastatic	Neoadjuvant	Control
N	132	52	19	15	46
Mean (\pm SD)	57.7 (\pm 13.9)	58.7 (\pm 12.0)	61.9 (\pm 9.8)	55.5 (\pm 16.6)	55.6 (\pm 13.7)
Median (range)	59.0 (0.0–83.0)	60.5 (27.0–80.0)	65.0 (44.0–75.0)	51.0 (29.0–79.0)	57.0 (26.0–83.0)
		Age			
		Grade			
1	10	10	-	-	
2	25	19	-	6	
3	30	16	8	6	
N/A	21	7	11	3	
		Stage			
I	15	15	-	-	
II	28	28	-	-	
III	9	9	-	-	
IV	19	-	19	-	
N/A	15	-	-	15	
		Lymphnode status			
Negative	27	24	-	3	
Positive	33	26	-	7	
N/A	26	2	19	5	
		Tumor size (before surgery)			
\leq 2	30	25	-	5	
>2 and \leq 6.5	33	26	-	7	
N/A	23	1	19	3	
		Metastatic sites			
Lung	12	-	12	-	
Skin	1	-	1	-	
Distantlymphnodes	5	-	5	-	
Pancreas	1	-	1	-	
Bone	9	-	9	-	
Liver	4	-	4	-	
Pleural	1	-	1	-	

Abbreviations: BrCa = breast cancer, SD = standard deviation, N/A = Non-Available or Non-Applicable.

Peripheral blood was collected in EDTA before treatment and was processed immediately for plasma isolation. In parallel, blood samples from 46 age-matched healthy female donors were included in our study (mean age: 55.65 (\pm SD) (\pm 13.7), median: 57.0 (range: 26.0–83.0)) (control group). All blood samples were centrifuged within 2 h twice at $2000\times g$ and then at $10,000\times g$ for 10 min and plasma was stored at $-80\text{ }^{\circ}\text{C}$ until further use.

4.2.2. ccfDNA Extraction

ccfDNA from plasma was extracted using the QIAamp DNA Blood Mini kit (Qiagen, MD, USA) according to manufacturer's instructions with some modifications. Specifically, DNA was eluted from 500 μL of plasma in 25 μL elution buffer and then stored in $-20\text{ }^{\circ}\text{C}$ until further use.

4.2.3. Sodium Bisulfite Conversion of ccfDNA

Bisulfite conversion was performed by EZ DNA Methylation-Gold™ Kit (ZYMO Research Co., Orange, CA, USA) as described by the manufacturer. During conversion, all unmethylated but not the methylated-cytosines of ccfDNA were converted to uracil. DNA was then eluted in 10 μL elution buffer and stored at $-80\text{ }^{\circ}\text{C}$ until further use. In each experiment, CpGenome Human methylated and nonmethylated DNA standards (Merck Millipore, Darmstadt, Germany) or H_2O were included as positive and negative controls respectively.

4.2.4. Quantitative Methylation-Specific PCR (qMSP)

Methylation of *ENPP2*s 1st Exon was analyzed by qMSP. *ENPP2* primers (Table 8) for methylated sequences specifically designed to contain cg02534163 were checked using Oligo 7 software and obtained from Eurofins, Genomic (Louisville, USA). A methylation-independent assay with non-CpG including primers for the β -actin gene (*ACTB*) was used in order to verify DNA quality and to normalize results. Specificity and cross-reactivity of methylated primers were evaluated by using SB-converted methylated and non-methylated DNA standards. Analytical sensitivity of qMSP assays was evaluated by using serial dilutions of SB-converted methylated and nonmethylated DNA standards (100%, 50%, 10%, 1%, 0%). The assay efficiency (expressed as $E = 10^{-1}/\text{slope}^{-1}$) was evaluated by using serial dilutions of the SB-converted methylated DNA standards in H_2O (100–0.01 ng). The results were calculated using the Rotor-Gene 6000 Series Software 1.7 (Qiagen). The analysis was performed according to the RQ sample (Relative Quantification) = $2^{-\Delta\Delta\text{CT}}$ method. Specifically, $\Delta\Delta\text{CT}$ values were generated for each target after normalization by *ACTB* values and using 100% methylation as calibrator.

Table 8. Primer sequences, annealing temperatures, and genomic locations used for qMSP assays.

GENE	Primer Sequence (5'–3')	Annealing Temperature ($^{\circ}\text{C}$)	Product Length	Genomic Loci
<i>ENPP2</i>	MET F: CGTTTTTTTATTGATACGATTGGAACGA	60	117bp	Chr8: 120650976–120651092 (+1 strand)
	MET R: CAAAACCT- CAAAACAAT- ACACTCCG- TAA			
<i>ACTB</i>	F: TGGTGATG- GAGGAGGTT- TAGTAAG R: AAC- CAATAAAACC- TACTCCTCCC	60	134bp	chr7: 5558705–5558838 (–1 strand)

Abbreviations: MET: methylated, F: forward, R: reverse.

In the *ENPP2* qMSP assay, a high linear correlation was found between the dilution ratios, analytical sensitivity of 0.01%, and efficiency 99%. Curves are presented in Supplementary Figure S2A. Additionally, analytical sensitivity of ACTB was 0.1% and efficiency was 96% (Supplementary Figure S2B).

4.2.5. Statistical Analysis

The Kolmogorov–Smirnov test was used to check for normality in distribution and the chi-squared test was used for comparison between discrete variables. One-way ANOVA test that was followed by Bonferroni post-hoc or Kruskal–Wallis test was applied to compare continuous variables between subgroups. In case of binary variables, t-test or Mann–Whitney test were also applied. Pearson or Spearman correlation was used for comparison between two continuous variables. Statistical significance was placed at p -value $< 5 \times 10^{-2}$. Statistical analysis was conducted with the IBM SPSS 19.0 statistical software (IBM Corp. 2010. IBM SPSS Statistics for Windows, Version 19.0. Armonk, NY, USA).

5. Conclusions

Presented data demonstrate *ENPP2* promoter hypermethylation in BrCa tissues associated with decreased expression, suggesting epigenetic regulation of its expression. Methylation events are correlated to BrCa progression and metastatic potential. In addition, we demonstrate that the *ENPP2* methylation assessed in liquid biopsy could offer a minimally invasive approach in early diagnosis and monitoring upon prospective evaluation. Our data introduce *ENPP2* methylation as a putative biomarker in BrCa.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms23073717/s1>.

Author Contributions: Conceptualization, M.P. and E.C.; methodology, M.P., E.C., D.F. and I.B.; software, M.P., D.F. and M.K.; validation, M.P., E.C. and A.D.; formal analysis, M.P., E.C., D.F. and A.D.; investigation, M.P., E.C., D.F. and A.D.; resources, E.C.; data curation, M.P., E.C., D.F., A.D., V.A. and N.X.; writing—original draft preparation, M.P., A.D. and D.F.; writing—review and editing, M.P., E.C., D.F., A.D., M.K., V.A., N.X. and I.B.; visualization, M.P. and D.F.; supervision, E.C., V.A. and N.X.; project administration, M.P. and E.C.; funding acquisition, E.C. All authors have read and agreed to the published version of the manuscript.

Funding: The research project was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “1st Call for H.F.R.I. Research Projects to Support Faculty Members & Researchers and the Procurement of High-Cost Research Equipment Grant” (Project Number: 1955).

Institutional Review Board Statement: The study was approved by the Scientific Board of the University General Hospital of Evros (PGNE), following assessment by Ethics Committee (decision 663/08.08.16), and was conducted.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Harbeck, N.; Gnant, M. Breast cancer. *Lancet* **2017**, *389*, 1134–1150. [[CrossRef](#)]
2. Akram, M.; Iqbal, M.; Daniyal, M.; Khan, A.U. Awareness and current knowledge of breast cancer. *Biol. Res.* **2017**, *50*, 33. [[CrossRef](#)] [[PubMed](#)]
3. Luque-Bolivar, A.; Villegas, V.E.; Rondón-Lagos, M. Resistance and Overcoming Resistance in Breast Cancer. *Breast Cancer Dove Med. Press* **2020**, *12*, 211–229. [[CrossRef](#)] [[PubMed](#)]
4. Teo, K.; Valerie, G.B. The role and therapeutic potential of the autotaxin—Lysophosphatidate signalling axis in breast cancer. *Biochem. J.* **2014**, *463*, 157–165. [[CrossRef](#)] [[PubMed](#)]
5. Benesch, M.G.K.; Tang, X.; Brindley, D.N. Autotaxin and Breast Cancer: Towards Overcoming Treatment Barriers and Sequelae. *Cancers* **2020**, *12*, 374. [[CrossRef](#)] [[PubMed](#)]

6. Magkrioti, C.; Galaris, A.; Kanellopoulou, P.; Stylianaki, E.A.; Kaffe, E.; Aidinis, V. Autotaxin and chronic inflammatory diseases. *J. Autoimmun.* **2019**, *104*, 102327. [[CrossRef](#)] [[PubMed](#)]
7. Choi, J.W.; Herr, D.R.; Noguchi, K.; Yung, Y.C.; Lee, C.W.; Mutoh, T.; Lin, M.E.; Teo, S.T.; Park, K.E.; Mosley, A.N.; et al. LPA receptors: Subtypes and biological actions. *Annu. Rev. Pharmacol. Toxicol.* **2010**, *50*, 157–186. [[CrossRef](#)]
8. Xu, Y.; Fang, X.J.; Casey, G.; Mills, G.B. Lysophospholipids activate ovarian and breast cancer cells. *Biochem. J.* **1995**, *309 Pt 3*, 933–940. [[CrossRef](#)]
9. Yang, S.Y.; Lee, J.; Park, C.G.; Kim, S.; Hong, S.; Chung, H.C.; Min, S.K.; Han, J.W.; Lee, H.W.; Lee, H.Y. Expression of autotaxin (NPP-2) is closely linked to invasiveness of breast cancer cells. *Clin. Exp. Metastasis* **2002**, *19*, 603–608. [[CrossRef](#)]
10. Tang, X.; Benesch, M.G.K.; Brindley, D.N. Role of the autotaxin–lysophosphatidate axis in the development of resistance to cancer therapy. *Biochim. Biophys. Acta BBA Mol. Cell Biol. Lipids* **2020**, *1865*, 158716. [[CrossRef](#)]
11. David, M.; Wannecq, E.; Descotes, F.; Jansen, S.; Deux, B.; Ribeiro, J.; Serre, C.-M.; Grès, S.; Bendriss-Vermare, N.; Bollen, M.; et al. Cancer Cell Expression of Autotaxin Controls Bone Metastasis Formation in Mouse through Lysophosphatidic Acid-Dependent Activation of Osteoclasts. *PLoS ONE* **2010**, *5*, e9741. [[CrossRef](#)]
12. Boucharaba, A.; Serre, C.-M.; Grès, S.; Saulnier-Blache, J.S.; Bordet, J.-C.; Guglielmi, J.; Clézardin, P.; Peyruchaud, O. Platelet-derived lysophosphatidic acid supports the progression of osteolytic bone metastases in breast cancer. *J. Clin. Investig.* **2004**, *114*, 1714–1725. [[CrossRef](#)] [[PubMed](#)]
13. Lim, D.H.K.; Maher, E.R. DNA methylation: A form of epigenetic control of gene expression. *Obstet. Gynaecol.* **2010**, *12*, 37–42. [[CrossRef](#)]
14. Panagopoulou, M.; Karaglani, M.; Balgkouranidou, I.; Bizioti, E.; Koukaki, T.; Karamitrousis, E.; Nena, E.; Tsamardinos, I.; Kolios, G.; Lianidou, E.; et al. Circulating cell-free DNA in breast cancer: Size profiling, levels, and methylation patterns lead to prognostic and predictive classifiers. *Oncogene* **2019**, *38*, 3387–3401. [[CrossRef](#)]
15. Panagopoulou, M.; Esteller, M.; Chatzaki, E. Circulating Cell-Free DNA in Breast Cancer: Searching for Hidden Information towards Precision Medicine. *Cancers* **2021**, *13*, 728. [[CrossRef](#)]
16. Panagopoulou, M.; Cheretaki, A.; Karaglani, M.; Balgkouranidou, I.; Bizioti, E.; Amarantidis, K.; Xenidis, N.; Kakolyris, S.; Baritaki, S.; Chatzaki, E. Methylation Status of Corticotropin-Releasing Factor (CRF) Receptor Genes in Colorectal Cancer. *J. Clin. Med.* **2021**, *10*, 2680. [[CrossRef](#)]
17. Panagopoulou, M.; Lambropoulou, M.; Balgkouranidou, I.; Nena, E.; Karaglani, M.; Nicolaidou, C.; Asimaki, A.; Konstantinidis, T.; Constantinidis, T.C.; Kolios, G.; et al. Gene promoter methylation and protein expression of BRMS1 in uterine cervix in relation to high-risk human papilloma virus infection and cancer. *Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med.* **2017**, *39*, 1010428317697557. [[CrossRef](#)]
18. Panagopoulou, M.; Fanidis, D.; Aidinis, V.; Chatzaki, E. ENPP2 Methylation in Health and Cancer. *Int. J. Mol. Sci.* **2021**, *22*, 11958. [[CrossRef](#)]
19. Brenet, F.; Moh, M.; Funk, P.; Feierstein, E.; Viale, A.J.; Socci, N.D.; Scandura, J.M. DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS ONE* **2011**, *6*, e14524. [[CrossRef](#)]
20. Wagner, J.R.; Busche, S.; Ge, B.; Kwan, T.; Pastinen, T.; Blanchette, M. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome Biol.* **2014**, *15*, R37. [[CrossRef](#)]
21. Brindley, D.N.; Tang, X.; Meng, G.; Benesch, M.G.K. Role of Adipose Tissue-Derived Autotaxin, Lysophosphatidate Signaling, and Inflammation in the Progression and Treatment of Breast Cancer. *Int. J. Mol. Sci.* **2020**, *21*, 5938. [[CrossRef](#)] [[PubMed](#)]
22. Parris, T.Z.; Kovács, A.; Hajizadeh, S.; Nemes, S.; Semaan, M.; Levin, M.; Karlsson, P.; Helou, K. Frequent MYC coamplification and DNA hypomethylation of multiple genes on 8q in 8p11-p12-amplified breast carcinomas. *Oncogenesis* **2014**, *3*, e95. [[CrossRef](#)] [[PubMed](#)]
23. Liu, X.; Peng, Y.; Wang, J. Integrative analysis of DNA methylation and gene expression profiles identified potential breast cancer-specific diagnostic markers. *Biosci. Rep.* **2020**, *40*, BSR20201053. [[CrossRef](#)] [[PubMed](#)]
24. Panagopoulou, M.; Karaglani, M.; Manolopoulos, V.G.; Iliopoulos, I.; Tsamardinos, I.; Chatzaki, E. Deciphering the Methylation Landscape in Breast Cancer: Diagnostic and Prognostic Biosignatures through Automated Machine Learning. *Cancers* **2021**, *13*, 1677. [[CrossRef](#)] [[PubMed](#)]
25. Ivan, J.; Patricia, G.; Agustriawan, D. In silico study of cancer stage-specific DNA methylation pattern in White breast cancer patients based on TCGA dataset. *Comput. Biol. Chem.* **2021**, *92*, 107498. [[CrossRef](#)]
26. Popnikolov, N.K.; Dalwadi, B.H.; Thomas, J.D.; Johannes, G.J.; Imagawa, W.T. Association of autotaxin and lysophosphatidic acid receptor 3 with aggressiveness of human breast carcinoma. *Tumour Biol. J. Int. Soc. Oncodev. Biol. Med.* **2012**, *33*, 2237–2243. [[CrossRef](#)] [[PubMed](#)]
27. Panagopoulou, M.; Karaglani, M.; Balgkouranidou, I.; Pantazi, C.; Kolios, G.; Kakolyris, S.; Chatzaki, E. Circulating cell-free DNA release in vitro: Kinetics, size profiling, and cancer-related gene methylation. *Cell. Physiol.* **2019**, *234*, 14079–14089. [[CrossRef](#)]
28. Wang, S.-C.; Liao, L.-M.; Ansar, M.; Lin, S.-Y.; Hsu, W.-W.; Su, C.-M.; Chung, Y.-M.; Liu, C.-C.; Hung, C.-S.; Lin, R.-K. Automatic Detection of the Circulating Cell-Free Methylated DNA Pattern of GCM2, ITPRIPL1 and CCDC181 for Detection of Early Breast Cancer and Surgical Treatment Response. *Cancers* **2021**, *13*, 1375. [[CrossRef](#)]
29. Wang, Z.; Jensen, M.A.; Zenklusen, J.C. A Practical Guide to The Cancer Genome Atlas (TCGA). *Methods Mol. Biol.* **2016**, *1418*, 111–141. [[CrossRef](#)]

30. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210. [[CrossRef](#)]
31. Jeschke, J.; Bizet, M.; Desmedt, C.; Calonne, E.; Dedeurwaerder, S.; Garaud, S.; Koch, A.; Larsimont, D.; Salgado, R.; Van den Eynden, G.; et al. DNA methylation-based immune response signature improves patient diagnosis in multiple cancers. *J. Clin. Investig.* **2017**, *127*, 3090–3102. [[CrossRef](#)] [[PubMed](#)]
32. Johnson, K.C.; Houseman, E.A.; King, J.E.; Christensen, B.C. Normal breast tissue DNA methylation differences at regulatory elements are associated with the cancer risk factor age. *Breast Cancer Res. BCR* **2017**, *19*, 81. [[CrossRef](#)] [[PubMed](#)]
33. Orozco, J.I.J.; Knijnenburg, T.A.; Manughian-Peter, A.O.; Salomon, M.P.; Barkhoudarian, G.; Jalas, J.R.; Wilmott, J.S.; Hothi, P.; Wang, X.; Takasumi, Y.; et al. Epigenetic profiling for the molecular classification of metastatic brain tumors. *Nat. Commun.* **2018**, *9*, 4627. [[CrossRef](#)] [[PubMed](#)]
34. Moss, J.; Magenheimer, J.; Neiman, D.; Zemmour, H.; Loyfer, N.; Korach, A.; Samet, Y.; Maoz, M.; Druid, H.; Arner, P.; et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat. Commun.* **2018**, *9*, 5068. [[CrossRef](#)] [[PubMed](#)]
35. Müller, F.; Scherer, M.; Assenov, Y.; Lutsik, P.; Walter, J.; Lengauer, T.; Bock, C. RnBeads 2.0: Comprehensive analysis of DNA methylation data. *Genome Biol.* **2019**, *20*, 55. [[CrossRef](#)] [[PubMed](#)]
36. Colaprico, A.; Silva, T.C.; Olsen, C.; Garofano, L.; Cava, C.; Garolini, D.; Sabedot, T.S.; Malta, T.M.; Pagnotta, S.M.; Castiglioni, I.; et al. TCGAAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **2016**, *44*, e71. [[CrossRef](#)]
37. Chandrashekar, D.S.; Bashel, B.; Balasubramanya, S.A.H.; Creighton, C.J.; Ponce-Rodriguez, I.; Chakravarthi, B.; Varambally, S. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* **2017**, *19*, 649–658. [[CrossRef](#)]

Obesity Reshapes the Microbial Population Structure along the Gut-Liver-Lung Axis in Mice

Η παχυσαρκία επαναδιαμορφώνει την δομή του μικροβιακού πληθυσμού κατά μήκος του άξονα εντέρου-ήπατος-πνεύμονα στα ποντίκια

Το μικροβίωμα αναδύεται ως μείζον παίκτης της ιστικής ομοιοστάσης υπό υγιείς αλλά και παθολογικές συνθήκες. Δυσβίωση του εντερικού μικροβιώματος συσχετίζεται με διάφορες αυτοφλεγμονώδεις και μεταβολικές ασθένειες. Δίαιτες πλούσιες σε λιπαρά και η επακόλουθη παχυσαρκία είναι γνωστό πως επηρεάζουν την πολυπλοκότητα και την ποικιλομορφία του μικροβιώματος και συνεπώς την παθοφυσιολογία του οργανισμού. Επιπροσθέτως έχει προταθεί η ύπαρξη ενός μικροβιακού άξονα μεταξύ εντέρου και ήπατος, άξονας που μπορεί να εκτείνεται και στους πνεύμονες. Σε αυτό το πλαίσιο, συγκρίναμε συστηματικώς το μικροβίωμα του εντέρου, του ήπατος και του πνεύμονα ποντικών οι οποίοι τρέφονταν με δίαιτα πλούσια σε λιπαρά με αυτό αδερφών ζώων που λάμβαναν την δίαιτα αναφοράς. Διεξάγαμε αλληλούχιση επτά υπερμεταβλητών περιοχών του 16S ριβοσωμικού RNA (rRNA) για να εξετάσουμε την μικροβιακή ποικιλότητα στους ιστούς ενδιαφέροντος. Σύγκριση των τοπικών μικροβιωμάτων υπέδειξε πως ο πνευμονικός ιστός φέρει το λιγότερο ποικιλόμορφο μικροβίωμα υπό ομοιοστατικές συνθήκες, ενώ το μικροβίωμα του υγιούς ήπατος ομοιάζει με αυτό του εντέρου. Η παχυσαρκία αύξησε την μικροβιακή πολυπλοκότητα και στους τρεις ιστούς, με την πνευμονική μικροβιακή ποικιλότητα να υφίσταται τις μεγαλύτερες αλλαγές. Επίσης, η παχύσαρκια προώθησε την επέκταση των Firmicutes κατά μήκος του άξονα εντέρου-συκωτιού-πνεύμονα, ενώ παράλληλα ανέδειξε το γένος *Staphylococcus* ως έναν πιθανό παθολογικό σύνδεσμο μεταξύ παχυσαρκίας και συστημικής παθοφυσιολογίας, ιδιαιτέρως δε των πνευμόνων.



Article

Obesity Reshapes the Microbial Population Structure along the Gut-Liver-Lung Axis in Mice

Apostolos Galaris ^{1,†}, Dionysios Fanidis ^{1,†}, Elli-Anna Stylianaki ¹, Vaggelis Harokopos ²,
Alexandra-Styliani Kalantzi ¹, Panagiotis Moulos ², Antigone S. Dimas ¹, Pantelis Hatzis ²
and Vassilis Aidinis ^{1,*}

¹ Institute of Bioinnovation, Biomedical Sciences Research Center Alexander Fleming, 16672 Athens, Greece; galaris@fleming.gr (A.G.); fanidis@fleming.gr (D.F.); stylianaki@fleming.gr (E.-A.S.); kalantzi@fleming.gr (A.-S.K.); dimas@fleming.gr (A.S.D.)

² Institute for Fundamental Biomedical Research, Biomedical Sciences Research Center Alexander Fleming, 16672 Athens, Greece; harokopos@fleming.gr (V.H.); moulos@fleming.gr (P.M.); hatzis@fleming.gr (P.H.)

* Correspondence: v.aidinis@fleming.gr

† These authors contributed equally to this work.

Abstract: The microbiome is emerging as a major player in tissue homeostasis in health and disease. Gut microbiome dysbiosis correlates with several autoimmune and metabolic diseases, while high-fat diets and ensuing obesity are known to affect the complexity and diversity of the microbiome, thus modulating pathophysiology. Moreover, the existence of a gut-liver microbial axis has been proposed, which may extend to the lung. In this context, we systematically compared the microbiomes of the gut, liver, and lung of mice fed a high-fat diet to those of littermates fed a matched control diet. We carried out deep sequencing of seven hypervariable regions of the 16S rRNA microbial gene to examine microbial diversity in the tissues of interest. Comparison of the local microbiomes indicated that lung tissue has the least diverse microbiome under healthy conditions, while microbial diversity in the healthy liver clustered closer to the gut. Obesity increased microbial complexity in all three tissues, with lung microbial diversity being the most modified. Obesity promoted the expansion of Firmicutes along the gut-liver-lung axis, highlighting staphylococcus as a possible pathologic link between obesity and systemic pathophysiology, especially in the lungs.

Keywords: obesity; high-fat diet; microbiome; 16S rRNA; gut; liver; lung; comparative analysis; firmicutes; staphylococcus



Citation: Galaris, A.; Fanidis, D.; Stylianaki, E.-A.; Harokopos, V.; Kalantzi, A.-S.; Moulos, P.; Dimas, A.S.; Hatzis, P.; Aidinis, V. Obesity Reshapes the Microbial Population Structure along the Gut-Liver-Lung Axis in Mice. *Biomedicines* **2022**, *10*, 494. <https://doi.org/10.3390/biomedicines10020494>

Academic Editor:
Eugenia Bezirtzoglou

Received: 3 January 2022

Accepted: 17 February 2022

Published: 19 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The microbiome, the sum of commensal, symbiotic, and pathogenic organisms that populate animal bodies, is increasingly recognized as a major player in tissue homeostasis in health and disease [1], modulating a variety of host functions, including immunity and inflammation [2], as well as energy homeostasis and metabolism [3]. Changes in microbial population structure and the ensuing local or systemic effects can be induced by different environmental factors, most notably exposure to antibiotics and dietary changes, while the efficacy of various medications has been suggested to correlate with microbiome perturbations and vice versa [1].

Most microorganisms reside within the intestine. It is well established that the gut microbiome participates in multiple homeostatic functions essential for the host, including nutrient absorption and education of the immune system. Alterations in the composition and complexity of microbiomes can harm the health of an organism. Such alterations lead to dysbiosis and have been associated with autoimmune and metabolic diseases, mostly via secreted microbial metabolites [4]. Non-alcoholic fatty liver disease (NAFLD) has been linked to dysbiosis [5,6], highlighting a connection between gut microbiota and the liver, referred to as the gut-liver axis [7]. The gut and liver are in direct contact

through the biliary tract and the portal vein, and disturbances in gut barrier functions result in an increased influx of bacteria and their metabolites to the liver [7]. Moreover, a gut-brain microbial axis has also been proposed [8], suggesting that the gut microbiome may also affect distant organs. In the same vein, a gut-lung axis has also been proposed, and dysbiosis has been associated most notably with asthma and chronic obstructive pulmonary disease (COPD) [9]. It is noteworthy that elder individuals become more susceptible to these respiratory diseases, while the process of aging has also been associated with microbial dysbiosis [10]. Moreover, the lungs—believed until recently to be sterile—have been found to contain their own bacterial flora, which is deregulated in disease states [11]. Specific bacterial species have recently been associated with disease status in COPD patients [12] and with mortality in idiopathic pulmonary fibrosis (IPF) patients [13]. Liver functions, including endotoxin and bacterial clearance, have been suggested as critical determinants of lung pathophysiology in acute respiratory distress syndrome (ARDS) [14], while the contribution of a liver-lung axis has also been proposed in alcohol-induced liver diseases [15]. Moreover, a fiber-rich diet has been shown to confer reduced risk for COPD, possibly through metabolic, liver-mediated effects on innate immunity. This points to a potential gut-liver-lung axis [16] that may further involve metabolic regulation of microbial dysbiosis. However, the gut-liver-lung axis is thought to be imposed mainly via microbial metabolites. In that respect, the actual microbial populations of the different organs along with the full gamut of metabolites and their possible similarities or differences have not been fully elucidated.

Obesity and obesity-related metabolic disorders are linked to lipid biosynthetic pathways in the liver and have been associated with the composition of the human gut microbiome [4,17]. To date, multiple studies are starting to highlight the important contribution of the microbiome to human health, but there remains a notable heterogeneity in published results. Moreover, much less is known about the effects of obesity on the microbial composition of the liver and lungs or along the gut-liver-lung axis. In this context, we examined the bacterial composition along the gut-liver-lung axis upon high-fat diet (HFD)-induced obesity in wild type (wt) C57Bl6 mice, based on 16 rRNA gene (V2-4, V6-9) sequencing.

2. Materials and Methods

2.1. Animals

All mice (four per dietary regime) were bred under specific-pathogen-free (SPF) conditions at the local animal facility at 20–22 °C, 55 ± 5% humidity, and a 12-h light/dark cycle; water and food were provided ad libitum. All experiments on mice were in line with ARRIVE guidelines and were approved by the Veterinary Service and Fishery Department of the local governmental prefecture (#740336), following the positive opinion of the Institutional Protocol Evaluation Committee BSRC Alexander Fleming. Experimental animals were on an HFD (60% fat VHFD, D12492i, Research diets, New Brunswick, New Jersey, USA), and their littermate control animals were fed the corresponding matched control diet (CD) (10% fat, D12450J, Research diets, New Brunswick, New Jersey, USA) for 16 weeks starting after their ablactation. All mice were co-housed continuously throughout the experiment, taking all necessary precautions for cross-contamination.

2.2. Tissue Sampling

Mice were anesthetized using a xylazine/ketamine/atropine mixture (10 mg/100 mg/0.05 mg/kg body weight respectively) and then euthanized by a gradual supply of carbon dioxide. Small intestines were isolated; the first ~1 cm right after stomach was discarded, and samples of ~50 mg were collected and immediately transferred into liquid nitrogen. No washing of the intestine was performed. Next, perfusion with PBS was performed, as previously described [18], and liver and lung samples were isolated and immediately transferred into liquid nitrogen. All tissue samples were stored at −80 °C until processing.

2.3. Histopathology

Right lung tissues and medial liver lobes were fixed using 10% neutral buffered formalin. The gut was fixed using a fixation buffer containing 50% Ethanol and 5% acetic acid. All tissues were placed in paraffin. 4 µm sections were prepared and stained with hematoxylin/eosin (H&E) standard protocols. Tissue imaging was performed using a Nikon Eclipse E800 microscope (Nikon Corp., Shinagawa-ku, Japan) attached to a Q Imaging EXI Aqua digital camera, using the Q-Capture Pro software (v7.0, QImaging, Surrey, BC, Canada).

2.4. Plasma Sampling

Plasma was collected from every subject as previously described (Barbayianni, Ninou, et al. 2018). In brief, after euthanasia of the animals, the blood was collected through the inferior vena cava, and EDTA was added to a final concentration of 10%. The samples were then centrifuged for 20 min at 2000 g at 4 °C. Plasma was transferred and stored at 4 °C until biochemical analysis was performed using a Beckman Coulter AU480 Clinical Chemistry Analyzer based at the BSRC 'Alexander Fleming' phenotyping facility for the estimation of Alanine Transaminase (ALT) (OSR6107, Beckman Coulter, Brea, CA, USA) and Aspartate Transaminase (AST) (OSR6109, Beckman Coulter, Brea, CA, USA) levels.

2.5. DNA Extraction and 16S Library Preparation

Total genomic DNA was extracted from approximately 50 mg of tissue using the DNeasy[®] Blood & Tissue Kit (Cat. Nos. 69504 and 69506, Qiagen, Hilden, Germany) following the manufacturer's suggested protocol. 2–8 µL (~100–250 ng) of genomic DNA from each sample was used with the Ion 16S[™] Metagenomics Kit (A26216, ThermoFisher Scientific, Waltham, MA, USA) to amplify the 16S hypervariable regions. The kit used includes two sets of primers targeting V2, V4, V8— And V3, V6–7, and V9 regions, respectively. After sample purification, DNA libraries were prepared with the Ion Plus Fragment Library Kit (4471252, ThermoFisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions. The libraries were pooled and sequenced on an Ion Proton[™] System. Briefly, the genomic DNA from each sample was divided into two parts, and each was amplified with a different pool of primers needed to cover the 16S hypervariable regions. Given that distinct hypervariable regions may contribute in varying degrees to the identification of bacterial species, especially for low-level taxa [19], we amplified seven of nine hypervariable regions (V2–4, V6–9), thus enabling broad-range identification of bacterial populations. Following amplification, the two reactions were pooled, and all samples were further purified with Agencourt AMPure XP Beads (A63881, Beckman Coulter, Brea, CA, USA). Sample concentration was measured using the Qubit dsDNA HS Assay Kit (Q32851 ThermoFisher Scientific, Waltham, MA, USA), and approximately 50 ng of amplified DNA was used to prepare DNA libraries with the Ion Plus Fragment Library Kit (4471252, ThermoFisher Scientific, Waltham, MA, USA). End repair was followed by adaptor ligation and nick repair. After purification, libraries were amplified and further purified. Each library's quality and quantity was assessed on a Bioanalyzer using the DNA High Sensitivity Kit reagents and protocol (5067-4626, Agilent Technologies, Santa Clara, CA, USA). Quantified libraries were pooled to a final concentration of 7 pM. The pools were then processed, templated, and enriched on an Ion Proton One Touch system. Templating was performed using the Ion PI[™] Hi-Q[™] OT2 200 Kit (A26433, ThermoFisher Scientific, Waltham, MA, USA), and sequencing was performed using the Ion PI[™] Hi-Q[™] Sequencing 200 Kit on Ion Proton PI[™] V3 chips (A26771, ThermoFisher Scientific, Waltham, MA, USA) according to commercially available protocols. The Ion Proton[™] System [20] was used for high-throughput sequencing according to the manufacturer's instructions. The above process is illustrated in Figure S1A.

2.6. 16S rRNA Gene Sequencing Data Analysis

Fastq files were quality controlled using DADA2 functions [21] in order to trim 14b read left ends, as suggested for IonTorrent data. Reads under 50 bp were subsequently removed along with those sequences with more than four expected errors. Trimmed reads were aligned against the human and mouse genomes using the FastQ Screen tool (v0.14.01) [22] to detect contaminant sequences. No microbial reads were filtered from the initial raw fastq files, which were pooled per sampled tissue and diet prior to quality control. Filtration and trimming were performed as before, prior to denoising and chimera removal using DADA2 (Figure S2; Table S1). Processed files were assigned to amplicon sequence variants (ASVs). Taxonomy assignment was performed using the SILVA database [23], followed by gene copy number (GCN) normalization. For GCN correction, abundance values were divided by each taxon's 16S gene copy number as recorded in rrnDB (v5.7 NCBI) [24] using an in-house script. Taxa with no records in rrnDB were not considered for downstream analysis. The above process is depicted in Figure S1B.

2.7. Statistical Analysis

Statistical analysis was performed using the Prism 6 software (GraphPad, San Diego, CA, USA), as specifically indicated in the text and the corresponding figure legends.

3. Results

3.1. High Fat Diet Induces Obesity and NAFLD

To examine the effect of obesity on the microbiome and possible interconnections of the gut-liver axis with the lung, we fed wt C57Bl6 mice with a non-toxic HFD for 16 weeks, starting right after their ab lactation, at four weeks of age. Littermate mice fed a matched CD were used as controls.

As expected, mice fed with HFD gained more weight than their littermate controls (Figure 1A). Obese mice presented with elevated Alanine Transaminase (ALT) levels in their plasma (Figure 1B) and decreased AST/ALT (AST: Aspartate Transaminase) ratios (Figure 1C), indicating liver damage as well as lipid deposition in the liver (Figure 1D), both reminiscent of NAFLD. Lipid deposition was also observed in the gut but not in the lung (Figure 1D).

3.2. Obesity Increases Microbial Complexity along the Gut-Liver-Lung Axis

Quality controlled, denoised fastq files were pooled per tissue and diet and were assigned to amplicon sequence variants (ASVs). Following abundance level correction for differing 16S gene copy numbers between taxa (GCN correction), we detected in total 29 phyla, 59 classes, 130 orders, 227 families, and 585 species (Figure S3A).

To reveal patterns of microbiota composition in each gut, liver, and lung, we calculated bacterial diversity in terms of microbial richness (observed number of ASVs) and biodiversity (Shannon's index). No differences were observed in host tissue 18S rRNA gene levels detected by RT-PCR and used as a negative loading control (Figure S3B). As shown in Figure 2A, obesity resulted in higher microbial richness compared to equivalent controls in all three tissues studied, an observation consistent with the higher numbers of microbial taxa assigned upon HFD (Figure S3C). We report that although the lung is the tissue with the least diverse microbiome under healthy conditions, its microbial diversity is affected the most by HFD-driven obesity and reaches levels similar to those recorded for the other two tissues under these conditions (Figure 2B).

To identify overall similarities in taxonomic composition between tissues (β -diversity), we calculated the Aitchison distance of tissue microbiomes under healthy and obese conditions. This distance metric was selected in order to take into consideration the compositional nature of 16S rRNA gene sequencing data [19]. As shown in Figure 2C, we observe that HFD-driven obesity shifts the composition of the lung microbiome closer to that of the liver. Moreover, HFD-driven obesity results in a relative increase in phyla and families that are shared across the examined tissues (Figure 2D,E; Tables S2 and S3).

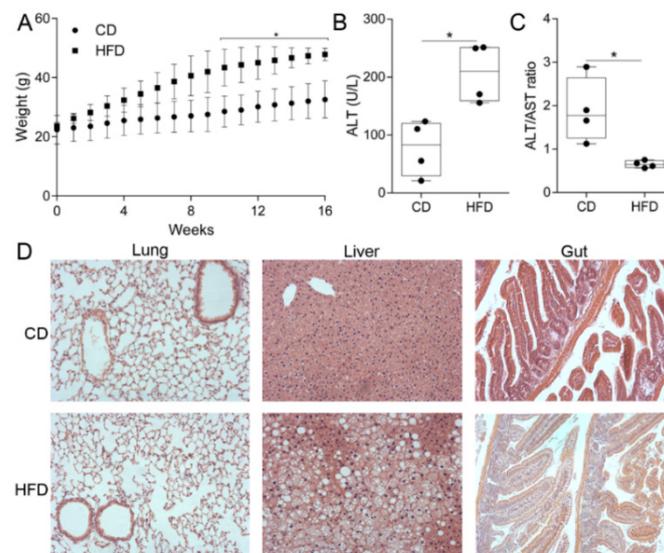


Figure 1. Mice fed with a high-fat diet (HFD) developed obesity and non-alcoholic liver fatty liver disease (NAFLD). Mice fed with HFD for 16 weeks presented with (A) statistically significant higher body weight from the tenth week of HFD onwards, (B) elevated ALT levels, and (C) decreased AST/ALT ratio in plasma after 16 weeks of HFD. (D) Representative images from the histopathology (hematoxylin & eosin staining) of gut, liver, and lung tissues, illustrating lipid deposition in the liver and gut after 16 weeks of HFD. Hematoxylin stains cell nuclei (purple) and eosin stains the extracellular matrix (pink). The “bubbles” appearing in HDF liver and gut samples are lipid droplets. Statistical significance was assessed through the Friedman test followed by pairwise Mann-Whitney tests (in A) and Mann-Whitney tests (in B and C); * $p < 0.05$ was considered significant.

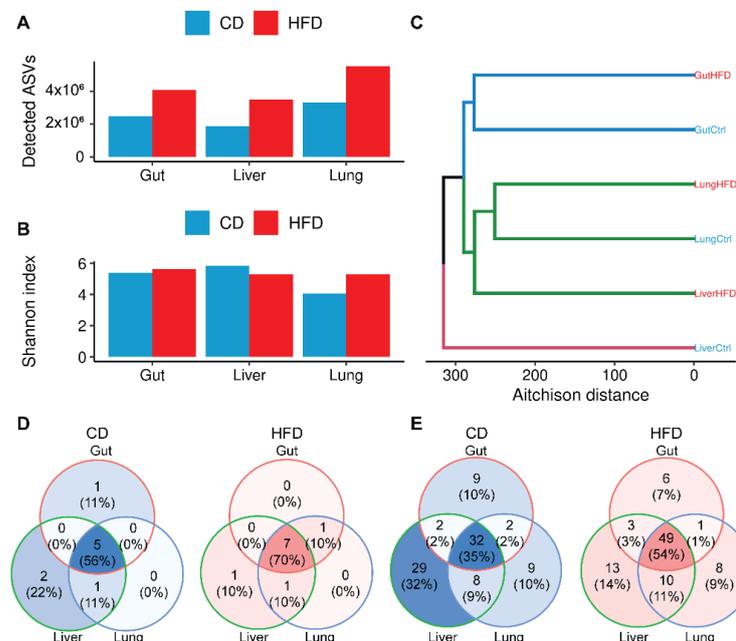


Figure 2. HFD-driven obesity triggers changes in the microbiome composition of all tissues examined. (A) HFD-driven obesity results in a greater number of detected Amplicon Sequence Variants (ASVs) in all tissues. (B) α -diversity per sample and diet. Shannon’s index was used to evaluate sample biodiversity per tissue and diet. (C) Similarity of samples per tissue and diet as described by β -diversity. Aitchison distance was used to account for the compositional nature of 16S rRNA sequencing data. (D) Venn diagrams of common phyla or (E) families upon CD or HFD in gut, liver, and lung.

3.3. Obesity Favors Firmicutes and, Most Notably, *Staphylococcus* Expansion in All Tissues

Given detected phyla, our data indicate that gut, liver, and lung share the same top-level taxa: Proteobacteria, Actinobacteria, Bacteroidetes, Cyanobacteria, and Firmicutes (Figure 3). Under CD and based on relative abundance, the gut microbiome is more closely related to that of the liver (Figure 3A). Upon HFD administration, this balance is disturbed, and the liver microbiome shifts closer to the lung (Figure 3B). This is also indicated by the recorded β -diversity (Figure 2C) and because more taxa are shared across the liver and lung at different taxonomic levels (Figure 2D,E).

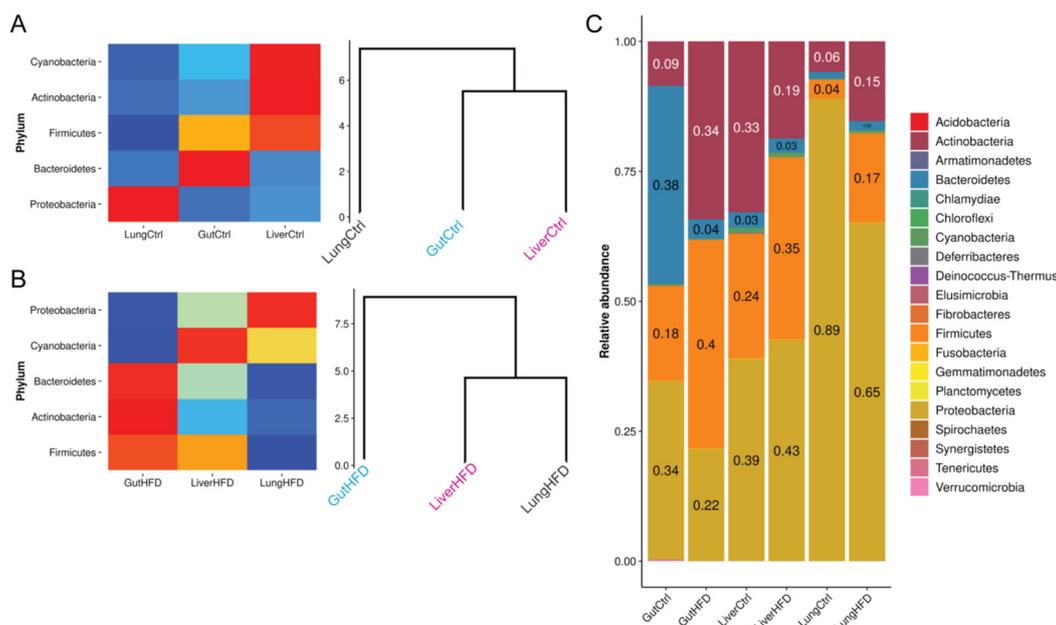


Figure 3. Relative abundance of phyla detected in gut, liver, and lung of control and HFD-fed mice. Heatmap and respective dendrogram of non-zero abundance phyla under CD (A) and HFD (B). (C) Relative abundance of all detected phyla. Relative abundance calculations for all panels were based on GCN values. For panels A and B, only phyla with an abundance greater than zero after value rounding to two decimal places were considered. Heatmaps are scaled per phylum. Manhattan distance was used to perform hierarchical clustering of the tissues with complete linkage in panels A and B.

In CD liver and lung tissue, we report that Proteobacteria are the most abundant bacteria, followed by Actinobacteria and Firmicutes (Figure 3C). This composition is slightly different in the CD gut, with the most abundant Bacteroidetes being followed by Proteobacteria and Firmicutes. Upon HFD, liver and lung are occupied, in order of abundance, by Proteobacteria, Firmicutes, and Actinobacteria. On the other hand, the gut is mainly populated by Firmicutes and Actinobacteria, followed by Proteobacteria. Notably, our findings show that HFD-driven obesity favors expansion of Firmicutes in all three tissues studied.

Deeper examination at the taxonomic level revealed that the Firmicutes families that expanded under HFD-driven obesity in all tissues examined were Staphylococcaceae, Streptococcaceae, and Peptoniphilaceae (Figure 4A); Pasteurellaceae family (Proteobacteria) expanded as well. Concerning detected genera, two, three, and eleven genera were detected within the Streptococcaceae, Staphylococcaceae, and Peptoniphilaceae taxa, respectively (Table S4). Among them, and under HFD-driven obesity, *Staphylococcus* has an increasing tendency in all three tissues relative to CD conditions (Figure 4B).

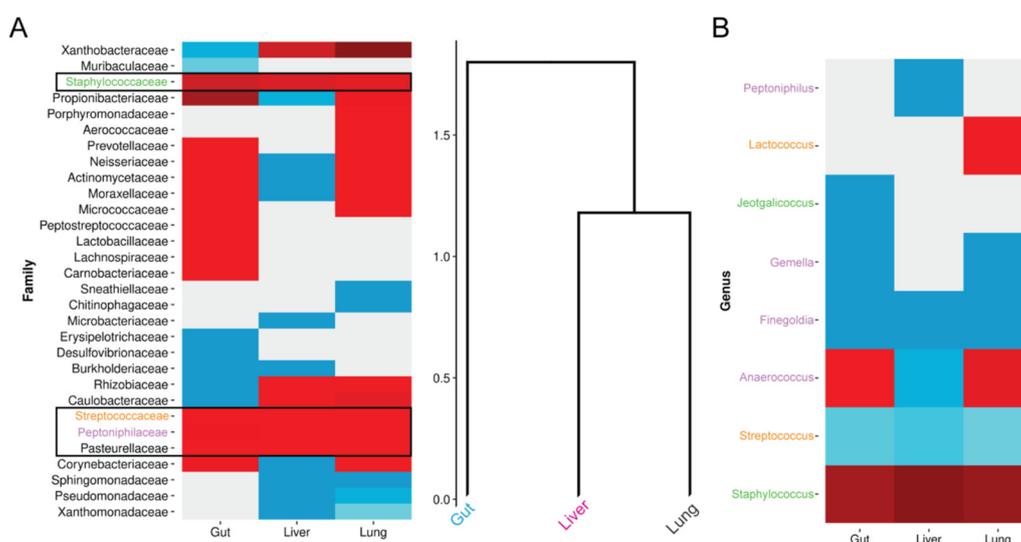


Figure 4. HFD consistently affects specific microbial families and genera. (A) Heatmap and tissue dendrogram of HFD-to-CD differences in family relative abundance (B) Heatmap of HFD-to-CD difference in genera relative abundance. For both panels, only those taxa with HFD-to-CD relative abundance differences other than zero in at least one tissue after rounding to two decimal places were considered. Heatmaps are scaled per taxon. Manhattan distance was used to perform hierarchical clustering of tissues with complete linkage. Relative abundance was calculated based on GCN abundance values. Families and respective genera are coded with the same color in both panels.

A notable point that we raise given our findings is that the three genera most consistently affected by HFD (Staphylococcus, Streptococcus, and Finegoldia) have been shown to produce superantigens (SAG), potent immunomodulators that are produced by microorganisms (Figure 4B). Given that SAGs are produced by specific, mostly bacterial species, we searched UniProt to track them in our dataset. A search of the database with the UniProt keyword ‘Superantigen’ revealed several UniProt-KB and Uni-Prot-TrEMBL entries, which, when intersected with those of our 16S amplicon sequencing, led to the detection of six SAG-related species (Table S5), of which four belong to genera whose diversity was affected upon administration of HFD in at least one of tissue. This points to the existence of a potential bacterial-driven pathogenic mechanism, which should be further explored.

4. Discussion

In this report, we examined the microbial composition of the guts, livers, and lungs of mice fed a high-fat diet (HFD) compared to littermates fed a matched control diet (CD). Towards this goal, we employed sequencing of seven (out of nine) 16S rRNA gene hypervariable regions. Obesity was shown to increase microbial complexity along the gut-liver-lung axis, promoting the expansion of Firmicutes, especially of Staphylococcus.

To induce obesity in mice, we utilized a non-toxic (not methionine- or choline-deficient) HFD containing 20% protein, 20% kcal fat, and 60% kcal carbohydrate with a total of 20% kcal energy density (5.21 kcal/g). As a result, the mice gained weight (unlike with toxic diets), with lipids being deposited in both liver and gut, but not in lung tissue (Figure 1).

The technique of 16S rRNA gene amplicon sequencing has revolutionized microbiomics [25]. Although most reported sequencing protocols examine one or two hypervariable regions [25], we selected seven hot spots (V2-V4 and V6-9) for amplification, given that each of these nine sub-regions has a distinct potential for distinguishing between microbial taxa, especially for lower taxonomic levels [19]. This selection also serves to avoid skewing of taxa distributions, as, for example, V1-V2 present poor performance in assigning sequences to the phylum of Proteobacteria [26], which are underrepresented in most studies. Additionally, amplicon sequence variants (ASVs) were preferred to operational taxonomic units (OTUs), given that OTUs represent certain bacterial taxa inadequately [27] and can

systematically polarize diversity metrics due to reference incompleteness [28]. Furthermore, we performed GCN correction to normalize microbial abundance levels for copy numbers.

In healthy CD mice, we report that all three tissues examined (gut, liver, and lung) were populated chiefly by four phyla: Proteobacteria, Actinobacteria, Bacteroidetes, and Firmicutes (Figure 3C), as previously reported for the gut in mice [29,30] and humans [31,32]. Interestingly, Cyanobacteria were also detected in all three tissues consistent with previous observations of the phylum in mouse lung [33], lower respiratory tract [34], and gastrointestinal canal [35]. Gut microbiome diversity is known to increase with age. In humans, it stabilizes at the age of three and is largely dominated by Bacteroidetes, Firmicutes, and Proteobacteria [36,37]. The liver is the first organ to encounter gut-derived bacteria upon intestinal barrier dysfunction through the portal vein, as well as bacteria from systemic infections via the hepatic artery, which are then removed from circulation through hepatic filtering [38,39]. Bacterial clearance by the liver has been suggested to occur in a dual-track mode: rapidly via Kupffer cell scavenger receptors and through a slower process involving different immune system mechanisms, making the translocated microorganisms available for the induction of adaptive immunity [38,39]. Given the above mechanisms, it is not surprising that bacterial rDNA was detected, to the best of our knowledge for the first time, in the healthy livers of mice (Figures 2–4). Our findings likely portray a snapshot of the process of bacterial clearance by the liver or remnants of degraded bacteria. Microbial diversity in the healthy liver clustered closer to that of the gut (C and Figure 4A), pointing to a larger contribution of portal vein-derived gut-leaked microbes than previously thought. This mechanism has been proposed to contribute to the worsening of human liver diseases [40]. Importantly, bacterial rDNA was recently detected in the liver of healthy human individuals [41]. Concerning lung tissue, which until recently was believed to be sterile, studies in mice have reported colonization with Proteobacteria, Firmicutes, and Bacteroidetes [42]. These phyla are also reported to be the most prevalent in the human adult lung [43,44]. We report that the mouse lung microbiome was dominated mainly by Proteobacteria, Firmicutes, Bacteroidetes, and Actinobacteria under both CD and HFD (Figure 3C), although α -diversity levels showed that the lung was the organ/tissue with the least diverse microbiome (Figure 2B).

Obesity is a complex disease characterized by extensive lipid deposition throughout the body and increases the risk for multiple diseases [45,46]. The gut microbiome contributes to the pathophysiology of obesity [47], and the 'obese' microbiome has been suggested to harvest energy from nutrients with increased capacity [48]. Gut microbiota have been shown to regulate body fat content in mice [11] and are regarded as a putative target for obesity treatments [47]. As shown in the present study, HFD-driven obesity suppresses Bacteroidetes in the gut of mice (Figure 3C), an effect also observed in aging mice [49]. Moreover, HFD-driven obesity was found to increase the Firmicutes to Bacteroidetes ratio in the gut (Figure 3C), expanding on previous studies in genetically engineered obese mice and obese humans [50,51]. An increase in the diversity of Firmicutes upon HFD-driven obesity was also observed, for the very first time, in the liver (Figure 3C). This observation may reflect gut microbiome leaking in the circulation and being cleared in the liver. However, a similar increase in Firmicutes was also recorded for lung tissue (Figure 3C), supporting the view that obesity stimulates systemic changes in microbiome composition, but also the existence of a gut-liver-lung axis. Such an inter-organ network maybe exist through different communication mechanisms that have been proposed in the past, including commensal microbe translocation [52] and chemical communication through bacterial metabolites (e.g., SCFAs) and tissue products [52,53].

Among Firmicutes, an abundance of *Staphylococcus* was characterized by an increasing tendency in all three tissues upon HFD-driven obesity (Figure 4). *Staphylococcus* is a Gram-positive, opportunistic pathogen that colonizes the skin as well as mucosal surfaces and may cause a range of infections in healthy and immunocompromised individuals, as well as in recovering postoperative patients. Obesity has been suggested to increase the risk of *Staphylococcus* colonization in humans, with obese individuals being more

susceptible to pneumonia, wound infections, bacteremia, and sepsis [54]. Increased abundance of Staphylococcus and Staphylococcaceae family members have also been reported in asthma [44], where obesity is a major risk factor and a disease modifier in children and adults [55]. The gut microbiome has been suggested to link obesity to asthma. Staphylococcus is a likely component of this link, with increased lung colonization by this microbe upon obesity. In cystic fibrosis (CF), *Staphylococcus aureus* is the second most commonly isolated pathogen from the airways of patients. The increasing prevalence of obesity in CF patients is associated with further impairment of lung function [56]. Lung Staphylococcus burden has also been found to be increased in chronic hypersensitivity pneumonitis (CHP) [57], has been associated with progression of idiopathic pulmonary fibrosis (IPF) [58], and a Staphylococcus pro-apoptotic peptide has been correlated with acute exacerbations of IPF [59]. Therefore, as the gut microbiome has been linked with that of the liver [40] and the lung [55] in different disease contexts and obesity, Staphylococcus may be a potential pathogenic link between the three organs under divergence from a steady state.

Furthermore, HFD-induced obesity was shown to affect bacterial species, mainly Staphylococcus, capable of producing superantigens (SAGs; Table S5), which are potent immunostimulatory molecules [60]. Chronic exposure to *S. aureus* SAG toxic shock syndrome toxin-1 (TSST-1) has been shown to facilitate the development of diabetic complications in rabbits [61], while TSST-1 has also been shown to stimulate cytokine production from adipocytes [62], thus possibly contributing to the low-grade systemic inflammation associated with obesity and diabetes. Moreover, staphylococcal SAG enterotoxin B (SEB) was shown to cause interstitial pneumonia in both autoimmune and non-autoimmune mice [63], although obesity is associated with decreased pneumonia risk and mortality, reflecting the ‘obesity paradox’.

The presented study is characterized by certain limitations. First, the pooled sample analysis design followed limits our results to a mostly descriptive nature. In addition, 16S rRNA gene amplicon sequencing data cannot be used directly for functional annotation. This, in combination with the study design, does not enable a deeper examination of proposed pathogenic mechanisms.

In conclusion, in addition to lipid deposition throughout the body and the triggering of NAFLD, obesity was shown to increase microbial complexity along the gut-liver-lung axis and, as a result, to possibly predispose mice to a series of metabolic diseases via increased abundance in Staphylococcus and other species.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biomedicines10020494/s1>, Figure S1: Schematic overview of experimental design and 16S rDNA sequencing and data analysis. Figure S2: Data quality control and error rate estimation; Figure S3: Obesity increased microbial complexity along the gut-liver-lung axis. Table S1: Quality control metrics before (grey-colored cells) and after (white-colored cells) human and bacteria sequences removal; Table S2. Important taxa as presented in Figure 2D,E.; Table S3: Number of shared taxa between any pairwise combination of experimental conditions; Table S4: Detected genera belonging to the Streptococcaceae, Staphylococcaceae, Peptoniphilaceae, and Pasteurellaceae families. Red marked are those affected by a high-fat diet in at least one tissue as defined by HFD-to-control relative abundance difference (non-zero difference); Table S5: Detected species related with superantigen proteins as recorded by UniProt database and/or respective literature.

Author Contributions: Conceptualization, V.A.; methodology, A.G., E.-A.S., V.H., P.M., P.H. and V.A.; in silico data analysis, D.F., A.-S.K., P.M. and A.S.D.; data curation, A.G., D.F., V.H., A.-S.K., P.M., A.S.D., P.H. and V.A.; writing—original draft preparation, A.G. and D.F.; writing—review and editing, P.M., P.H. and V.A. All authors have read and agreed to the published version of the manuscript.

Funding: A.G. was supported by a fellowship from the Hellenic Foundation for Research & Innovation (# 789 to A.G.). D.F. was supported by European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship, and Innovation, under the call Research—Create—Innovate (T1EDK-0049 to V.A.).

Institutional Review Board Statement: The animal study protocol was approved by the Veterinary Service and Fishery Department of the local governmental prefecture (#740336), following approval of the Institutional Protocol Evaluation Committee of BSRC Alexander Fleming.

Informed Consent Statement: Not applicable.

Data Availability Statement: Processed data presented in this study are openly available in Zenodo at DOI: <https://doi.org/10.5281/zenodo.6141436>.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Cho, I.; Blaser, M.J. The human microbiome: At the interface of health and disease. *Nat. Rev. Genet.* **2012**, *13*, 260–270. [[CrossRef](#)] [[PubMed](#)]
2. Belkaid, Y.; Hand, T.W. Role of the Microbiota in Immunity and Inflammation. *Cell* **2014**, *157*, 121–141. [[CrossRef](#)] [[PubMed](#)]
3. Cani, P.D.; Van Hul, M.; Lefort, C.; Depommier, C.; Rastelli, M.; Everard, A. Microbial regulation of organismal energy homeostasis. *Nat. Metab.* **2019**, *1*, 34–46. [[CrossRef](#)] [[PubMed](#)]
4. Fan, Y.; Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* **2021**, *19*, 55–71. [[CrossRef](#)]
5. Jiang, W.; Wu, N.; Wang, X.; Chi, Y.; Zhang, Y.; Qiu, X.; Hu, Y.; Li, J.; Liu, Y. Dysbiosis gut microbiota associated with inflammation and impaired mucosal immune function in intestine of humans with non-alcoholic fatty liver disease. *Sci. Rep.* **2015**, *5*, srep08096. [[CrossRef](#)] [[PubMed](#)]
6. Zhu, L.; Baker, S.S.; Gill, C.; Liu, W.; Alkhoury, R.; Baker, R.D.; Gill, S.R. Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH) patients: A connection between endogenous alcohol and NASH. *Hepatology* **2013**, *57*, 601–609. [[CrossRef](#)]
7. Tripathi, A.; Debelius, J.; Brenner, D.A.; Karin, M.; Looma, R.; Schnabl, B.; Knight, R. The gut–liver axis and the intersection with the microbiome. *Nat. Rev. Gastroenterol. Hepatol.* **2018**, *15*, 397–411. [[CrossRef](#)]
8. Morais, L.H.; Schreiber, H.L., IV; Mazmanian, S.K. The gut microbiota–brain axis in behaviour and brain disorders. *Nat. Rev. Microbiol.* **2021**, *19*, 241–255. [[CrossRef](#)]
9. Zhang, D.; Li, S.; Wang, N.; Tan, H.-Y.; Zhang, Z.; Feng, Y. The Cross-Talk Between Gut Microbiota and Lungs in Common Lung Diseases. *Front. Microbiol.* **2020**, *11*, 301. [[CrossRef](#)]
10. Saint-Criq, V.; Lugo-Villarino, G.; Thomas, M. Dysbiosis, malnutrition and enhanced gut-lung axis contribute to age-related respiratory diseases. *Ageing Res. Rev.* **2021**, *66*, 101235. [[CrossRef](#)]
11. Bäckhed, F.; Ding, H.; Wang, T.; Hooper, L.V.; Koh, G.Y.; Nagy, A.; Semenkovich, C.F.; Gordon, J.I. The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15718–15723. [[CrossRef](#)] [[PubMed](#)]
12. Bowerman, K.L.; Rehman, S.F.; Vaughan, A.; Lachner, N.; Budden, K.F.; Kim, R.Y.; Wood, D.L.A.; Gellatly, S.L.; Shukla, S.D.; Wood, L.G.; et al. Disease-associated gut microbiome and metabolome changes in patients with chronic obstructive pulmonary disease. *Nat. Commun.* **2020**, *11*, 1–15. [[CrossRef](#)] [[PubMed](#)]
13. Mackintosh, J.A.; Desai, S.R.; Adamali, H.; Patel, K.; Chua, F.; Devaraj, A.; Kouranos, V.; Kokosi, M.; Margaritopoulos, G.; Renzoni, E.; et al. In patients with idiopathic pulmonary fibrosis the presence of hiatus hernia is associated with disease progression and mortality. *Eur. Respir. J.* **2019**, *53*, 1802412. [[CrossRef](#)] [[PubMed](#)]
14. Herrero, R.; Sánchez, G.; Asensio, I.; López, E.; Ferruelo, A.; Vaquero, J.; Moreno, L.; de Lorenzo, A.; Bañares, R.; Lorente, J.A. Liver–lung interactions in acute respiratory distress syndrome. *Intensiv. Care Med.* **2020**, *8*, 1–13. [[CrossRef](#)] [[PubMed](#)]
15. Arteel, G.E. Liver–lung axes in alcohol-related liver disease. *Clin. Mol. Hepatol.* **2020**, *26*, 670–676. [[CrossRef](#)] [[PubMed](#)]
16. Young, R.P.; Hopkins, R.; Marsland, B.J. The Gut–Liver–Lung Axis. Modulation of the Innate Immune Response and Its Possible Role in Chronic Obstructive Pulmonary Disease. *Am. J. Respir. Cell Mol. Biol.* **2016**, *54*, 161–169. [[CrossRef](#)]
17. Stanislawski, M.A.; Dabelea, D.; Lange, L.A.; Wagner, B.D.; Lozupone, C.A. Gut microbiota phenotypes of obesity. *Biofilms Microbiomes* **2019**, *5*, 1–9. [[CrossRef](#)]
18. Barbayianni, I.; Ninou, I.; Tzouveleakis, A.; Aidinis, V. Bleomycin Revisited: A Direct Comparison of the Intratracheal Micro-Spraying and the Oropharyngeal Aspiration Routes of Bleomycin Administration in Mice. *Front. Med.* **2018**, *5*, 269. [[CrossRef](#)]
19. Gloor, G.B.; Macklaim, J.M.; Pawlowsky-Glahn, V.; Egozcue, J.J. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front. Microbiol.* **2017**, *8*, 2224. [[CrossRef](#)]
20. Rothberg, J.M.; Hinz, W.; Rearick, T.M.; Schultz, J.; Mileski, W.; Davey, M.; Leamon, J.H.; Johnson, K.; Milgrew, M.J.; Edwards, M.; et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **2011**, *475*, 348–352. [[CrossRef](#)]
21. Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Holmes, S.P. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **2016**, *13*, 581–583. [[CrossRef](#)] [[PubMed](#)]
22. Wingett, S.; Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* **2018**, *7*, 1338. [[CrossRef](#)] [[PubMed](#)]
23. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2013**, *41*, D590–D596. [[CrossRef](#)]

24. Stoddard, S.F.; Smith, B.J.; Hein, R.; Roller, B.R.; Schmidt, T.M. rrnDB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* **2015**, *43*, D593–D598. [[CrossRef](#)]
25. Bharti, R.; Grimm, D.G. Current challenges and best-practice protocols for microbiome analysis. *Briefings Bioinform.* **2021**, *22*, 178–193. [[CrossRef](#)] [[PubMed](#)]
26. Johnson, J.S.; Spakowicz, D.J.; Hong, B.-Y.; Petersen, L.M.; Demkowicz, P.; Chen, L.; Leopold, S.R.; Hanson, B.M.; Agresta, H.O.; Gerstein, M.; et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **2019**, *10*, 5029. [[CrossRef](#)] [[PubMed](#)]
27. Větrovský, T.; Baldrian, P. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLoS ONE* **2013**, *8*, e57923. [[CrossRef](#)] [[PubMed](#)]
28. Callahan, B.J.; McMurdie, P.J.; Holmes, S.P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **2017**, *11*, 2639–2643. [[CrossRef](#)]
29. Liu, C.; Zhou, N.; Du, M.-X.; Sun, Y.-T.; Wang, K.; Wang, Y.-J.; Li, D.-H.; Yu, H.-Y.; Song, Y.; Bai, B.-B.; et al. The Mouse Gut Microbial Biobank expands the coverage of cultured bacteria. *Nat. Commun.* **2020**, *11*, 1–12. [[CrossRef](#)]
30. Lagkouvardos, I.; Pukall, R.; Abt, B.; Foessel, B.U.; Meier-Kolthoff, J.P.; Kumar, N.; Bresciani, A.; Martínez, I.; Just, S.; Ziegler, C.; et al. Erratum: Corrigendum: The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nat. Microbiol.* **2016**, *1*, 16219. [[CrossRef](#)]
31. Mahowald, M.A.; Rey, F.E.; Seedorf, H.; Turnbaugh, P.J.; Fulton, R.S.; Wollam, A.; Shah, N.; Wang, C.; Magrini, V.; Wilson, R.K.; et al. Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 5859–5864. [[CrossRef](#)] [[PubMed](#)]
32. Turnbaugh, P.J.; Hamady, M.; Yatsunencko, T.; Cantarel, B.L.; Duncan, A.; Ley, R.E.; Sogin, M.L.; Jones, W.J.; Roe, B.A.; Affourtit, J.P.; et al. A core gut microbiome in obese and lean twins. *Nature* **2009**, *457*, 480–484. [[CrossRef](#)] [[PubMed](#)]
33. Barfod, K.K.; Roggenbuck, M.; Hansen, L.H.; Schjørring, S.; Larsen, S.T.; Sørensen, S.J.; Kroghfelt, K.A. The murine lung microbiome in relation to the intestinal and vaginal bacterial communities. *BMC Microbiol.* **2013**, *13*, 303. [[CrossRef](#)] [[PubMed](#)]
34. Zhang, R.; Chen, L.; Cao, L.; Li, K.-J.; Huang, Y.; Luan, X.-Q.; Li, G. Effects of smoking on the lower respiratory tract microbiome in mice. *Respir. Res.* **2018**, *19*, 253. [[CrossRef](#)]
35. Gu, S.; Chen, D.; Zhang, J.-N.; Lv, X.; Wang, K.; Duan, L.-P.; Nie, Y.; Wu, X.-L. Bacterial Community Mapping of the Mouse Gastrointestinal Tract. *PLoS ONE* **2013**, *8*, e74957. [[CrossRef](#)]
36. Yatsunencko, T.; Rey, F.E.; Manary, M.J.; Trehan, I.; Dominguez-Bello, M.G.; Contreras, M.; Magris, M.; Hidalgo, G.; Baldassano, R.N.; Anokhin, A.P.; et al. Human gut microbiome viewed across age and geography. *Nature* **2012**, *486*, 222–227. [[CrossRef](#)]
37. Rinninella, E.; Raoul, P.; Cintoni, M.; Franceschi, F.; Miggiano, G.A.D.; Gasbarrini, A.; Mele, M.C. What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. *Microorganisms* **2019**, *7*, 14. [[CrossRef](#)]
38. Broadley, S.P.; Plaumann, A.; Coletti, R.; Lehmann, C.; Wanisch, A.; Seidlmeier, A.; Esser, K.; Luo, S.; Rämer, P.C.; Massberg, S.; et al. Dual-Track Clearance of Circulating Bacteria Balances Rapid Restoration of Blood Sterility with Induction of Adaptive Immunity. *Cell Host Microbe* **2016**, *20*, 36–48. [[CrossRef](#)]
39. Zeng, Z.; Surewaard, B.G.; Wong, C.; Geoghegan, J.A.; Jenne, C.N.; Kubes, P. CRIg Functions as a Macrophage Pattern Recognition Receptor to Directly Bind and Capture Blood-Borne Gram-Positive Bacteria. *Cell Host Microbe* **2016**, *20*, 99–106. [[CrossRef](#)]
40. Iebba, V.; Guerrieri, F.; Di Gregorio, V.; Levrero, M.; Gagliardi, A.; Santangelo, F.; Sobolev, A.P.; Circi, S.; Giannelli, V.; Mannina, L.; et al. Combining amplicon sequencing and metabolomics in cirrhotic patients highlights distinctive microbiota features involved in bacterial translocation, systemic inflammation and hepatic encephalopathy. *Sci. Rep.* **2018**, *8*, 1–14. [[CrossRef](#)]
41. Suppli, M.P.; Bagger, J.I.; Lelouvier, B.; Broha, A.; Demant, M.; König, M.J.; Strandberg, C.; Lund, A.; Vilsbøll, T.; Knop, F.K. Hepatic microbiome in healthy lean and obese humans. *JHEP Rep.* **2021**, *3*, 100299. [[CrossRef](#)] [[PubMed](#)]
42. Gollwitzer, E.S.; Saglani, S.; Trompette, A.; Yadava, K.; Sherburn, R.; McCoy, K.D.; Nicod, L.P.; Lloyd, C.; Marsland, B.J. Lung microbiota promotes tolerance to allergens in neonates via PD-L1. *Nat. Med.* **2014**, *20*, 642–647. [[CrossRef](#)] [[PubMed](#)]
43. Erb-Downward, J.R.; Thompson, D.L.; Han, M.K.; Freeman, C.M.; McCloskey, L.; Schmidt, L.A.; Young, V.B.; Toews, G.B.; Curtis, J.L.; Sundaram, B.; et al. Analysis of the lung microbiome in the “healthy” smoker and in COPD. *PLoS ONE* **2011**, *6*, e16384. [[CrossRef](#)] [[PubMed](#)]
44. Hilty, M.; Burke, C.; Pedro, H.; Cardenas, P.; Bush, A.; Bossley, C.; Davies, J.; Ervine, A.; Poulter, L.; Pachter, L.; et al. Disordered microbial communities in asthmatic airways. *PLoS ONE* **2010**, *5*, e8578. [[CrossRef](#)] [[PubMed](#)]
45. Pugliese, G.; Liccardi, A.; Graziadio, C.; Barrea, L.; Muscogiuri, G.; Colao, A. Obesity and infectious diseases: Pathophysiology and epidemiology of a double pandemic condition. *Int. J. Obes.* **2022**, 1–17. [[CrossRef](#)] [[PubMed](#)]
46. González-Muniesa, P.; Martínez-González, M.-A.; Hu, F.B.; Després, J.-P.; Matsuzawa, Y.; Loos, R.J.F.; Moreno, L.A.; Bray, G.A.; Martínez, J.A. Obesity. *Nat. Rev. Dis. Prim.* **2017**, *3*, 1–18. [[CrossRef](#)]
47. Muscogiuri, G.; Cantone, E.; Cassarano, S.; Tuccinardi, D.; Barrea, L.; Savastano, S.; Colao, A. Gut microbiota: A new path to treat obesity. *Int. J. Obes. Suppl.* **2019**, *9*, 10–19. [[CrossRef](#)]
48. Turnbaugh, P.J.; Ley, R.E.; Mahowald, M.A.; Magrini, V.; Mardis, E.R.; Gordon, J.I. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **2006**, *444*, 1027–1031. [[CrossRef](#)]
49. Binyamin, D.; Werbner, N.; Nuriel-Ohayon, M.; Uzan, A.; Mor, H.; Abbas, A.; Ziv, O.; Teperino, R.; Gutman, R.; Koren, O. The aging mouse microbiome has obesogenic characteristics. *Genome Med.* **2020**, *12*, 1–9. [[CrossRef](#)]

50. Ley, R.E.; Turnbaugh, P.J.; Klein, S.; Gordon, J.I. Human gut microbes associated with obesity. *Nature* **2006**, *444*, 1022–1023. [[CrossRef](#)]
51. Ley, R.E.; Bäckhed, F.; Turnbaugh, P.; Lozupone, C.A.; Knight, R.D.; Gordon, J.I. Obesity alters gut microbial ecology. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 11070–11075. [[CrossRef](#)] [[PubMed](#)]
52. Wang, R.; Tang, R.; Li, B.; Ma, X.; Schnabl, B.; Tilg, H. Gut microbiome, liver immunology, and liver diseases. *Cell. Mol. Immunol.* **2021**, *18*, 4–17. [[CrossRef](#)] [[PubMed](#)]
53. Mathieu, E.; Escribano-Vazquez, U.; Descamps, D.; Cherbuy, C.; Langella, P.; Riffault, S.; Remot, A.; Thomas, M. Paradigms of Lung Microbiota Functions in Health and Disease, Particularly, in Asthma. *Front. Physiol.* **2018**, *9*, 1168. [[CrossRef](#)] [[PubMed](#)]
54. Befus, M.; Lowy, F.D.; Miko, B.A.; Mukherjee, D.V.; Herzig, C.T.A.; Larson, E.L. Obesity as a Determinant of Staphylococcus aureus Colonization Among Inmates in Maximum-Security Prisons in New York State. *Am. J. Epidemiol.* **2015**, *182*, 494–502. [[CrossRef](#)] [[PubMed](#)]
55. Peters, U.; Dixon, A.E.; Forno, E. Obesity and asthma. *J. Allergy Clin. Immunol.* **2018**, *141*, 1169–1179. [[CrossRef](#)] [[PubMed](#)]
56. Litvin, M.; Yoon, J.C.; Casella, J.L.; Blackman, S.M.; Brennan, A.L. Energy balance and obesity in individuals with cystic fibrosis. *J. Cyst. Fibros.* **2019**, *18*, S38–S47. [[CrossRef](#)] [[PubMed](#)]
57. Invernizzi, R.; Wu, B.; Barnett, J.; Ghai, P.; Kingston, S.; Hewitt, R.J.; Feary, J.; Li, Y.; Chua, F.; Wu, Z.; et al. The Respiratory Microbiome in Chronic Hypersensitivity Pneumonitis Is Distinct from That of Idiopathic Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* **2021**, *203*, 339–347. [[CrossRef](#)]
58. Han, M.K.; Zhou, Y.; Murray, S.; Tayob, N.; Noth, I.; Lama, V.N.; Moore, B.B.; White, E.S.; Flaherty, K.R.; Huffnagle, G.B.; et al. Lung microbiome and disease progression in idiopathic pulmonary fibrosis: An analysis of the COMET study. *Lancet Respir. Med.* **2014**, *2*, 548–556. [[CrossRef](#)]
59. D’Alessandro-Gabazza, C.N.; Kobayashi, T.; Yasuma, T.; Toda, M.; Kim, H.; Fujimoto, H.; Hataji, O.; Takeshita, A.; Nishihama, K.; Okano, T.; et al. A Staphylococcus pro-apoptotic peptide induces acute exacerbation of pulmonary fibrosis. *Nat. Commun.* **2020**, *11*, 1539. [[CrossRef](#)]
60. Fraser, J.D.; Proft, T. The bacterial superantigen and superantigen-like proteins. *Immunol. Rev.* **2008**, *225*, 226–243. [[CrossRef](#)]
61. Vu, B.G.; Stach, C.S.; Kulhankova, K.; Salgado-Pabon, W.; Klingelhutz, A.J.; Schlievert, P.M. Chronic Superantigen Exposure Induces Systemic Inflammation, Elevated Bloodstream Endotoxin, and Abnormal Glucose Tolerance in Rabbits: Possible Role in Diabetes. *mBio* **2015**, *6*, e02554-14. [[CrossRef](#)] [[PubMed](#)]
62. Vu, B.G.; Gourronc, F.A.; Bernlohr, D.A.; Schlievert, P.M.; Klingelhutz, A.J. Staphylococcal Superantigens Stimulate Immortalized Human Adipocytes to Produce Chemokines. *PLoS ONE* **2013**, *8*, e77988. [[CrossRef](#)] [[PubMed](#)]
63. Shinbori, T.; Matsuki, M.; Suga, M.; Kakimoto, K.; Ando, M. Induction of Interstitial Pneumonia in Autoimmune Mice by Intratracheal Administration of Superantigen Staphylococcal Enterotoxin B. *Cell. Immunol.* **1996**, *174*, 129–137. [[CrossRef](#)] [[PubMed](#)]

Increased lipocalin-2 expression in pulmonary inflammation and fibrosis

Αυξημένη έκφραση της lipocalin-2 στην πνευμονική φλεγμονή και ίνωση

Εισαγωγή: Η ιδιοπαθής πνευμονική ίνωση (idiopathic pulmonary fibrosis· IPF) είναι μία χρόνια προοδευτική διάμεση πνευμονοπάθεια με ζοφερή πρόγνωση. Οι υποκείμενοι παθολογικοί μηχανισμοί είναι ελάχιστα κατανοητοί οδηγώντας σε έλλειψη αποτελεσματικών θεραπειών. Εντούτοις, επανειλημμένη καταστροφή του πνευμονικού επιθηλίου θεωρείται κριτικής σημασίας για την έναρξη και την εξέλιξη της ασθένειας μέσω της έκκρισης διαλυτών παραγόντων. Οι τελευταίοι ενισχύουν την φλεγμονή, ενώ οδηγούν σε ενεργοποίηση ινοβλαστών και άφθονη εναπόθεση συστατικών της εξωκυτταρίας ουσίας. Η LCN2 είναι μια πρωτεΐνη που έχει προταθεί ως βιοδείκτης καταστροφής του ήπατος. Το ίδιο μόριο έχει αναφερθεί να τροποποιεί την έμφυτη ανοσία συμπεριλαμβανομένης της στρατολόγησης ουδετεροφίλων και να προστατεύει ενάντια σε βακτηριακές λοιμώξεις μέσω της απομόνωσης σιδήρου.

Μέθοδοι: *In silico* ανάλυση δημοσίως διαθέσιμων μεταγραφωμικών συνόλων δεδομένων· ELISA βρογχοκυψελιδικού υγρού ασθενών της IPF· επαγωγή πνευμονικής φλεγμονής και ίνωσης χρήσει BLM και επαγωγή οξέος πνευμονικού τραυματισμού χρήσει LPS σε ποντικούς· δοκιμασίες πνευμονικής λειτουργικότητας, ιστολογία, Q-RT-PCR, ανοσοαποτύπωση κατά Western και FACS ανάλυση στα παραπάνω ζωϊκά μοντέλα

Αποτελέσματα: Ανιχνεύθηκαν αυξημένα επίπεδα μεταγραφής της LCN2 σε πνευμονικό ιστό ασθενών της IPF, αλλαγή που συσχετίστηκε αρνητικά με λειτουργίες της αναπνοής. Ίδιες παρατηρήσεις έγιναν και για την πρωτεΐνη όπως αυτή ανιχνεύθηκε στο βρογχοκυψελιδικό υγρό κοόρτης ασθενών. Αυξημένη έκφραση της LCN2 εντοπίστηκε επίσης κατά την διάρκεια πνευμονικής φλεγμονής και ίνωσης στο πειραματικό μοντέλο της μπλεομυκίνης (BLM), με κορύφωση στην οξεία φάση, η οποία σχετίζεται με εισροή ουδετεροφίλων. Αύξηση παρατηρήθηκε και κατά την διάρκεια οξέος πνευμονικού τραυματισμού μετά από έκθεση σε LPS. Ανέλπιστα, και έχοντας υπόψιν τους περιορισμούς της παρούσας έρευνας, το ποντίκι *Lcn2^{-/-}* βρέθηκε να αναπτύσει πνευμονική φλεγμονή και ίνωση μετά από έκθεση σε BLM ή LPS, προβληματίζοντας σχετικά με τον κύριο παθολογικό ρόλο της *Lcn2* στα ποντίκια. Εντούτοις, και προτρέποντας για εκτενέστερες μελέτες, η LCN2 πληρεί τα κριτήρια για χρήση ως αναπληρωτής (surrogate) βιοδείκτης πνευμονικής φλεγμονής, αλλά και ως ένας πιθανός δείκτης επηρεασμένων πνευμονικών λειτουργιών.



OPEN ACCESS

EDITED BY

Hyungsoon Im,
Massachusetts General Hospital and Harvard
Medical School, United States

REVIEWED BY

Michael Adam O'Reilly,
University of Rochester, United States
Miho Jeong,
Massachusetts General Hospital, United States

*CORRESPONDENCE

Vassilis Aidinis
✉ V.Aidinis@Fleming.gr

RECEIVED 28 March 2023

ACCEPTED 07 August 2023

PUBLISHED 07 September 2023

CITATION

Galaris A, Fanidis D, Tsitoura E,
Kanellopoulou P, Barbayianni I, Ntatsoulis K,
Touloumi K, Gramenoudi S, Karampitsakos T,
Tzouvelekis A, Antoniou K and Aidinis V (2023)
Increased lipocalin-2 expression in pulmonary
inflammation and fibrosis.
Front. Med. 10:1195501.
doi: 10.3389/fmed.2023.1195501

COPYRIGHT

© 2023 Galaris, Fanidis, Tsitoura,
Kanellopoulou, Barbayianni, Ntatsoulis,
Touloumi, Gramenoudi, Karampitsakos,
Tzouvelekis, Antoniou and Aidinis. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted which
does not comply with these terms.

Increased lipocalin-2 expression in pulmonary inflammation and fibrosis

Apostolos Galaris¹, Dionysios Fanidis¹, Eliza Tsitoura²,
Paraskevi Kanellopoulou¹, Ilianna Barbayianni¹,
Konstantinos Ntatsoulis¹, Katerina Touloumi¹, Sofia Gramenoudi¹,
Theodoros Karampitsakos³, Argyrios Tzouvelekis³,
Katerina Antoniou² and Vassilis Aidinis^{1*}

¹Institute for Fundamental Biomedical Research, Biomedical Sciences Research Center Alexander Fleming, Athens, Greece, ²Department of Respiratory Medicine, School of Medicine, University of Crete, Heraklion, Greece, ³Department of Respiratory Medicine, School of Medicine, University of Patras, Patras, Greece

Introduction: Idiopathic Pulmonary Fibrosis (IPF) is a chronic, progressive interstitial lung disease with dismal prognosis. The underlying pathogenic mechanisms are poorly understood, resulting in a lack of effective treatments. However, recurrent epithelial damage is considered critical for disease initiation and perpetuation, via the secretion of soluble factors that amplify inflammation and lead to fibroblast activation and exuberant deposition of ECM components. Lipocalin-2 (LCN2) is a neutrophil gelatinase-associated lipocalin (NGAL) that has been suggested as a biomarker of kidney damage. LCN2 has been reported to modulate innate immunity, including the recruitment of neutrophils, and to protect against bacterial infections by sequestering iron.

Methods: *In silico* analysis of publicly available transcriptomic datasets; ELISAs on human IPF patients' bronchoalveolar lavage fluids (BALFs); bleomycin (BLM)-induced pulmonary inflammation and fibrosis and LPS-induced acute lung injury (ALI) in mice: pulmonary function tests, histology, Q-RT-PCR, western blot, and FACS analysis.

Results and discussion: Increased *LCN2* mRNA expression was detected in the lung tissue of IPF patients negatively correlating with respiratory functions, as also shown for BALF LCN2 protein levels in a cohort of IPF patients. Increased *Lcn2* expression was also detected upon BLM-induced pulmonary inflammation and fibrosis, especially at the acute phase correlating with neutrophilic infiltration, as well as upon LPS-induced ALI, an animal model characterized by neutrophilic infiltration. Surprisingly, and notwithstanding the limitations of the study and the observed trends, *Lcn2*^{-/-} mice were found to still develop BLM- or LPS-induced pulmonary inflammation and fibrosis, thus questioning a major pathogenic role for *Lcn2* in mice. However, LCN2 qualifies as a surrogate biomarker of pulmonary inflammation and a possible indicator of compromised pulmonary functions, urging for larger studies.

KEYWORDS

idiopathic pulmonary fibrosis (IPF), bleomycin (BLM), acute lung injury, transcriptomics, lipocalin-2 (LCN2)

Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive interstitial lung disease characterized by the exuberant deposition of collagens and other ECM components by lung fibroblasts, leading to the distortion of lung architecture and the impairment of respiratory functions. The underlying mechanisms of the disease are poorly understood, resulting in a

lack of effective treatments. However, epithelial damage is considered a key event initiating the pathogenesis of IPF, where repeated injury and/or abnormal repair of the epithelium trigger a cascade of signaling events that result in the recruitment and activation of immune cells, as well as the activation and accumulation of lung fibroblasts (1, 2).

Expression profiling of human IPF samples has been instrumental in the discovery of novel pathogenic genes and cellular pathways (3), that some were validated in animal models and some were translated into the clinic (4). In this context, we have recently developed Fibromine, a database and data mining tool, hosting all publicly available IPF transcriptomic (and proteomic) datasets (5), thus allowing the further exploitation of legacy data. Comparative analysis selected several hundred genes as differentially expressed in IPF, while an explainable machine learning phenotype classification algorithm prioritized 76 genes that include previously identified IPF expression hallmarks (e.g., Col1a1), IPF biomarkers (e.g., MMP7), as well as many genes previously shown to be involved in the pathophysiology of IPF (e.g., SPP1) (6). Among the novel, commonly identified deregulated genes in IPF was Lipocalin-2 (LCN2), also known as neutrophil gelatinase B-associated lipocalin (NGAL), as it was initially identified in neutrophilic granules in association with matrix metalloproteinase 9 (MMP9; gelatinase B) (7, 8). However, LCN2 secretion from other immune cells, as well as epithelial cells, has been reported (9, 10). LCN2 is considered an acute-phase protein, and increased LCN2 expression has been reported in different pathophysiological situations, including heart failure, kidney disease, and gut inflammation (10).

In the lung, increased LCN2 expression has been reported in subclinical pulmonary emphysema (11), chronic obstructive pulmonary disease (COPD) (12, 13), acute respiratory distress syndrome (ARDS) (14), as well as in patients with influenza A and SARS-CoV-2 virus infections (15). Not surprisingly, given their commonalities (16), higher LCN2 expression in bronchial epithelial cells of IPF patients has been also reported (17). Moreover, and more intriguingly, LCN2 has been suggested to mediate innate immune responses to bacterial infection by sequestering iron (18), whereas both iron homeostasis (19), as well as microbiome regulation (20), have been linked with IPF pathogenesis. Therefore, in this report, we investigated a possible role for LCN2 in pulmonary inflammation and fibrosis, by using *in silico* analysis of publicly available transcriptomic datasets, examination of LCN2 protein levels in IPF patients, as well as *in vivo* mouse models of pulmonary inflammation and fibrosis.

Materials and methods

Datasets

All analyzed bulk-sequencing datasets (Supplementary Table S1) were sourced from Fibromine (5). scRNA sequencing (scRNAseq) datasets used in the study are detailed in Supplementary Table S2.

TABLE 1 Demographics and clinical characteristics of IPF patients.

Characteristic	IPF (n = 26)
Age (yr) (Mean ± SD)	72.8 ± 7.3
Gender, n (%)	
Male	25 (96.1%)
Female	1 (3.9%)
Pulmonary function tests (mean ± SD)	
DLCO%	56.2 ± 19.4
FEV1/FVC%	85.4 ± 4.7
KCO	94.1 ± 21.5
Hematological analysis (%) (mean ± SD)	
Macrophages	83.1 ± 9.8
Lymphocytes	7.6 ± 7.6*
Polymorphonuclear	7.5 ± 6.4
Eosinophils	1.3 ± 1.61
BALF LCN2 (ng/mL) eosinophils	58.9 ± 52.3

*FVC%, Forced vital capacity percent predicted; DLCO%, Carbon monoxide diffusing capacity percent predicted; FEV1%, Forced expiratory volume in 1-s percent predicted.

Human patients

All studies were performed in accordance with the Declaration of Helsinki principles at the Department of Thoracic Medicine, University Hospital of Heraklion, and the demographics and clinical characteristics of the IPF patients can be found in Table 1. The diagnosis of IPF was based on ATS/ERS criteria or multidisciplinary discussion according to the Fleischer criteria (2, 21). Patients were anti-fibrotic naïve. All patients were evaluated with complete pulmonary function tests (PFTs) within 1 month of bronchoscopy. Lung volumes were measured using body plethysmography and the diffusion capacity (DLco, corrected for hemoglobin) using the single breath technique, and a computerized system (Jaeger 2.12; MasterLab, Würzburg, Germany). Patients were classified as non-smokers, current smokers, or former smokers (defined as having smoked a minimum of one cigarette a day for a minimum of 1 year, and stopping at least 6 months before presentation). All patients provided written informed consent. The study was approved by the Ethics Committees of the University Hospital of Heraklion (IRB numbers: 1045 and 17030).

Mice

All mice were bred at the animal facilities of the Alexander Fleming Biomedical Sciences Research Center under specific pathogen-free conditions. Mice were housed at 20–22°C, 55 ± 5% humidity, and a 12-h light–dark cycle; water and food were given *ad libitum*. Mice were bred and maintained in their respective genetic backgrounds for more than 10 generations. All experiments performed on mice for this project were in line with the ARRIVE guidelines and were approved by the Institutional Animal Ethical Committee (IAEC) of the Biomedical Sciences Research Center

“Alexander Fleming” (#373/375), as well as the Veterinary service and Fishery Department of the local governmental prefecture (#5508). Lipocalin-2 deficient mice (*Lcn2^{-/-}*) were procured from the Jackson Laboratory (#024630) and were maintained in a C57Bl6 genetic background for over 10 generations; genotyping was performed as previously published (18). Mice were humanely euthanized in a gradually filled CO₂ chamber.

BLM-induced pulmonary fibrosis

Pulmonary fibrosis was induced through the administration of bleomycin (BLM, 0.8 U/Kg of body weight; Nippon Kayaku) to anesthetized mice (intraperitoneal; ketamine/xylazine/atropine, 100/10/0.05 mg/Kg, respectively) via the oropharyngeal (OA) track, as previously described (22). In brief, mice were carefully placed on a plastic wall upon anesthesia. Their tongue was pulled out with forceps to get a better view of the trachea. The nares were blocked to force inhalation, and the bleomycin, diluted in normal saline (~50 µl for each mouse), was directly delivered to the oropharyngeal cavity using a conventional pipette tip. Normal saline was administered in the same way to littermate mice used as controls 3, 7, and 14 days after bleomycin (or saline) administration, at the peak of BLM-induced disease (which spontaneously resolves at d21).

Lipopolysaccharide (LPS)-induced acute lung injury (ALI)

The acute lung injury (ALI) model was performed using LPS delivered by inhalation, as previously described (23). In brief, bacterial lipopolysaccharides (LPS) from *Pseudomonas aeruginosa* (serotype 10, Sigma, St. Louis, MO, USA) were dissolved in normal saline at a concentration of 2 mg/ml. A total of 5 ml of this solution was administered into a chamber containing 5–7 mice via a custom-made nebulizer at an oxygen flow rate of 4 lt/min for 25 min. Normal saline was administered to the control mice. All measures were taken to minimize animal suffering; however, during the protocol, no anesthetics were used as no invasive or painful techniques were performed. After the induction of ALI, the condition of the animals was checked every 2 h during the light period. Mice were euthanized 24 h after the induction of ALI.

Respiratory functions

The respiratory functions were examined with FlexiVent (Scireq), following the manufacturer's instructions and as previously published (22).

Analyses of samples

Blood was collected through the portal vein and placed into tubes containing 0.5 M EDTA at a concentration of 10% v/v. Then, it was centrifuged for 20 min at 2,000 g at 4°C, and the plasma

was transferred in new siliconized tubes. BALF was obtained by lavaging the airways with 3 ml of normal saline using a cannula through the trachea (three times; 1 mL each). Then, BALF was centrifuged for 15 min at 1,200 g at 4°C. The first 1 ml of the BALF was transferred without the cells into a new siliconized tube. The other 2 ml were discarded; the cells were pooled and treated with GEYS solution for 10 min in ice. Then, they were centrifuged for 10 min at 1,200 g at 4°C, the suspension was discarded, the cell pellet was re-suspended in fresh PBS, and the cells were counted under an inverted microscope using a Neubauer chamber. The left lung lobe was cut and instantly transferred into liquid nitrogen for RNA and protein extraction. The remaining lobes were filled with formalin (143091.1214, AppliChem), to be later mounted into paraffin. Additionally, total protein concentration was estimated in the BALF using Bradford reagent (Cat.no.: 39222.03, SERVA) following the manufacturer's instructions.

Flow cytometry

Mice were euthanized under deep anesthesia followed by exsanguination. Then, BALF was collected via tracheotomy by injecting and slowly withdrawing 3 ml (3 times; 1 mL each) of phosphate-buffered saline (PBS). The cells were collected via 10 min centrifugation at 1,200 rpm at 4°C, and they were treated with 1 mL of Gey's Solution for 2–3 min. The Gey's Solution was removed after a 10 min centrifugation at 1,200 rpm at 4°C, and the cells were resuspended in 1XPBS/1 %FBS and counted manually under a reversed light microscope using an improved Neubauer hemacytometer according to common procedures. Next, the cells were centrifuged at 1,200 rpm for 10 min at 4°C. The cell pellets were resuspended in 50 µl blocking buffer (1XPBS with 1% FBS and 1:400 CD16/32) for 10 min. Then, 100 µL of PBS was added to each sample, and the cells were collected via 5 min centrifugation at 1200 rpm at 4°C. The cells were resuspended and stained in the desired concentrations of antibodies in 1XPBS + 1%FBS for 30 min. Then, 100 µL of PBS was added to each sample, and the cells were collected via 5 min centrifugation at 1200 rpm at 4°C. Finally, the cells were resuspended in 250 µl filtered PBS, and data were acquired on a BD FACSCanto TM II flow cytometer using BD FACSDiva software (BD Biosciences). The analyses of the RAW data were performed with the FlowJo software (TreeStar, Ashland, OR).

RNA extraction and real-time PCR

The upper half of the left lobe isolated from the animals was homogenized in 1 mL of Trizol (TR118, Molecular Research Center) followed by total RNA extraction according to the manufacturer's instructions. A total of 2 µg of total RNA were used for cDNA construction using M-MLV reverse transcriptase (28025-013, Invitrogen) according to the manufacturer's instructions. Real-time polymerase chain reaction (RT-PCR) was performed using SoFAst EvaGreen Supermix on a Bio-Rad CFX96 Touch™ real-time PCR detection system (Bio-Rad Laboratories Ltd, CA, USA). Values were normalized to β2-microglobulin (B2M) and the

primers used are: Lcn2 (F: 5'-GGG AAA TAT GCA CAG GTA TCC TC-3'; R: 5'-CAT GGC GAA CTG GTT GTA GTC-3') and B2M (F: 5'-TTC TGG TGC TTG TCT CAC TGA-3'; R: 5'-CAG TAT GTT CGG CTT CCC ATT C-3').

Protein extraction and Western blot analysis

The lower half of the left lobe was homogenized in 100 μ L of homemade RIPA cell lysis buffer (20 mM Tris-HCl pH = 7.5, 150 mM NaCl, 2 mM EDTA, 1 mM EGTA, 0.5% Sodium Deoxycholate, 0.1% SDS, 1% N-P40) containing protease inhibitor mixture (Cat. No: 11836170001, Roche) using a manual tissue grinder, and lysates were spun at 10,000 rpm for 10 min at 4°C. Protein concentration was estimated by Bradford reagent (Cat. no.: 39222.03, SERVA), and 10 μ g of total protein was prepared for immunoblotting in the final volume of 15 μ L. In detail, protein mixtures were incubated at 100°C for 5 min, and they were immediately spun and electrophoresed in SDS-PAGE gel. Proteins were then transferred onto nitrocellulose blotting membrane (GE10600002, Amersham, Germany), and the membranes were incubated in 1% BSA-1% Tween20 PBS in 1:1200 rabbit anti-mouse lipocalin-2 antibody (ab63929, Abcam) and 1:1,200 goat anti-mouse actin antibody (sc-1615, Santa Cruz Biotechnology) O/N at 4°C. The next day, the membranes were washed in 1% Tween20 PBS followed by incubation with 1:20,000 secondary antibodies (anti-rabbit: 925-68073, LI-COR; anti-goat: 925-32214, LI-COR) in 1% BSA-1% Tween20 PBS. The blot was visualized in an Odyssey DLx Imaging System (LI-COR).

Immunohistochemistry

Fixed lung tissues were mounted into paraffin; 4 μ m slices were cut and placed on slides. Then, hematoxylin/eosin (H&E) staining was performed as previously described (22). In brief, the slices were deparaffinized at 60°C for 2 h followed by xylene washes and hydrated in gradual ethanol concentrations. The slices were stained against Lcn2 (ab63929, Abcam) in 1:200 concentration. Peroxidase conjugated secondary antibody (4010-05, Southern Biotech) and DAB kit (SK-4100, Vector Laboratories, Inc.) were used to visualize Lcn2 in the lung tissue slices.

ELISA

LCN2 levels were estimated in human and murine BALF using a commercially available ELISA kit (EA100541, OriGene Technologies Inc.), according to the manufacturer's instructions.

In silico analyses

Differential gene expression analysis results produced during Fibromine creation (5) were used for volcano plot creation. Respective boxplots summarize LCN2 expression in terms of

log₂ fold change, while depicted datasets had a statistically significant difference between the compared groups (IPF_vs_Ctrl; Bleomycin_vs_Ctrl). Absolute fold change (FC) of at least 1.2 and FDR-corrected $p < 0.05$ were selected as thresholds for differential expression. The correlation of LCN2 expression values with those of spirometry measurements was examined using Spearman's correlation test. An absolute rho value of at least 0.5 was considered the threshold of a strong relationship, while a $p < 0.05$ was required for a relationship to be deemed significant. Visualizations were performed using packages ggplot2 (v.3.3.5) and ggrepel (v.0.9.1).

Single-cell RNA-seq data were found at GSE136831 (24), (GSE135893_ILD_annotated_fullsize.rds.gz) (25), as well as in the GitHub repositories (26, 27). All downstream described processes were completed with the R package Seurat (4.0.5) (28, 29).

For GSE136831, already filtered data were log normalized using a scaling factor of ten thousand (*NormalizeData*), and highly variable features (HVG) were retrieved (*FindVariableFeatures*) and scaled (*ScaleData*). Linear dimensionality reduction (PCA) (*RunPCA*) was followed by the creation of the closest neighborhood graph (*FindNeighbors*) using the first 7 principal components, as proposed by the median of all findPC methods output (30). Clusters were identified using Louvain clustering with a resolution of 1.3 (*FindClusters*). Cell typing information provided along with the count data was adopted. Non-linear dimensionality reduction was performed using Uniform Manifold Approximation and Projection (UMAP) (*RunUMAP*). As its name implies, UMAP is a non-linear method for reducing the dimensions of a dataset based on manifold calculation (31). Although not developed for scRNA-seq data *per se*, it is a method of choice for the analysis of such data yielding reproducible results in fast running times (32). Taking into consideration the same number of principal components as above results in a visualization very similar to that of the initial publication. Batch correction of any kind was not performed as proven unnecessary during the original data analysis.

From the GSE135893 object, IPF and control originating cells were maintained, while read counts were log normalized with a scaling factor of ten thousands before any downstream analysis (*NormalizeData* function). Mayr et al. lung dataset object was analyzed for IPF and control cells only, while barcodes that were assigned an "empty" cell type were removed. Log normalization with a 10,000 scaling was applied (*NormalizeData* function). Similarly, barcodes assigned a "NA" or "Low-Quality Cells" cell type were removed from the Strunz et al. whole lung dataset object before downstream analysis.

For all single-cell data differential expression analyses, the Wilcoxon rank-sum test was applied (*FindMarkers*), while an absolute FC of at least 1.2 and a Bonferroni-corrected $p < 0.05$ were set as significant thresholds.

Statistics

Statistical analysis was performed using the GraphPad Prism software (v8.0, GraphPad, San Diego, California, USA), as explicitly indicated in each figure legend.

Results

Increased *LCN2* expression in IPF patients negatively correlates with respiratory functions

To explore a possible involvement of *LCN2* in IPF, *LCN2* expression was interrogated in IPF transcriptomic datasets (Supplementary Table S1), sourced from Fibromine (www.fibromine.com), a database and data mining tool for target discovery in IPF (5). Using absolute fold change of at least 1.2 and FDR-corrected $p < 0.05$, widely accepted thresholds for the selection of differentially expressed genes, the expression of *LCN2* was found to be significantly increased in most datasets interrogating gene expression in the lungs of IPF patients in comparison with control individuals (Supplementary Table S1; Figure 1A). Indicatively, *LCN2* presented with a natural scale fold change of 2.3, 4.5, and 3.9 in three of the largest ones (Figure 1B; Supplementary Table S1). Importantly, *LCN2* expression negatively correlated with the respiratory functions (DLCO, FVC, and FEV1) of IPF patients in the same datasets (Figure 1C; Supplementary Figures S1A, B).

To examine the cell specificity of *LCN2* expression in fibrotic lungs, we re-analyzed data from three publicly available single-cell RNA seq (scRNAseq) datasets of human origin (Supplementary Table S2) (24–26). *LCN2* was found in all three data collections (Figure 1D; Supplementary Figures S1C, D), primarily expressed in epithelial cells, including goblet, ciliated, basal, club, and aberrant basaloid cells (Figures 1D, E; Supplementary Figures S1E, F). Comparing cell types between phenotypes (IPF and control), *LCN2* was found over-expressed mostly in alveolar type 1 and 2 cells (AT1 and AT2) (Figure 1F; Supplementary Figures S1G, H). However, *LCN2* expression from neutrophils, as shown in other pathological contexts summarized by the CellMarker2.0 database (33) (Supplementary Table S3), cannot be excluded, given the low representation of neutrophils in human IPF scRNAseq datasets.

To validate the *in silico* findings, we estimated *LCN2* levels in the bronchoalveolar lavage fluid of 26 IPF patients (Table 1), with a commercially available ELISA kit. As shown *in silico* for mRNA levels in the lung tissue of IPF patients (Figure 1), *LCN2* BALF levels of IPF patients negatively correlated with their respiratory functions (FEV1/FVC, TLCO, and KCO) (Table 1; Figures 2B, C).

Therefore, IPF is associated with increased *LCN2* expression, predominantly in pulmonary epithelial cells, negatively correlating with impaired lung functions.

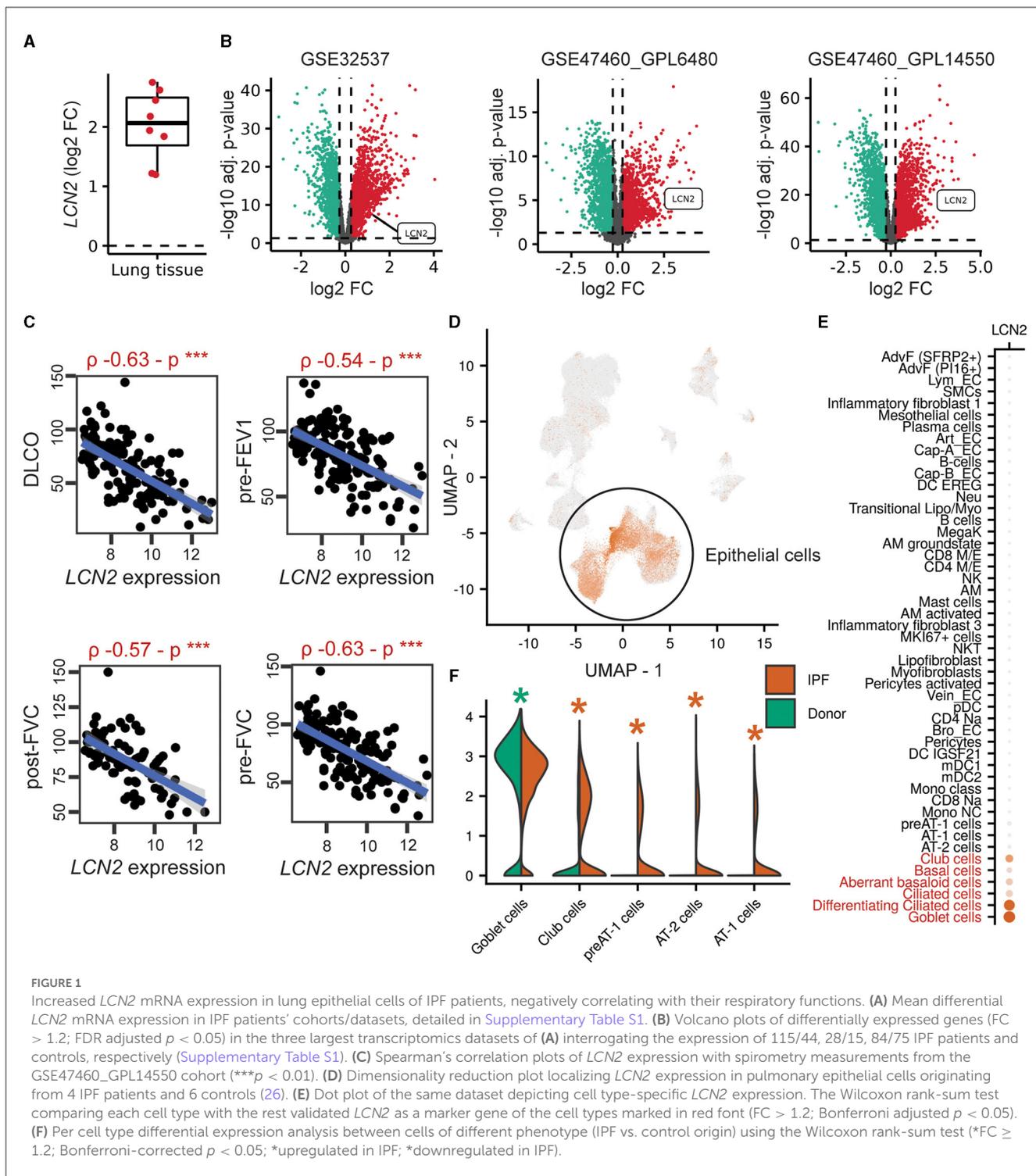
Increased *Lcn2* expression upon pulmonary inflammation and fibrosis in mice

To examine *Lcn2* expression in the lungs of mice post-bleomycin (BLM)-induced pulmonary inflammation and fibrosis, a widely used animal model of pulmonary fibrosis (4, 22, 34), we mined the relative transcriptomic datasets from Fibromine (Supplementary Table S1), as in IPF patients. *Lcn2* was found over-expressed in most datasets when comparing the fibrotic phase of the model to control samples (Supplementary Table S1; Figure 3A),

with indicative natural scale fold change scores of 5.4, 3, and 4.2 (Figure 3B). Moreover, re-analysis of a publicly available murine scRNAseq dataset (Supplementary Table S2) (27) indicated that, as in the human lung, *Lcn2* is highly expressed mainly by epithelial cells, as well as neutrophils (Figures 3C, D). More specifically, classical and activated AT2 cells, neutrophils, goblet, and activated mesothelial cells, as well as lymphatic endothelial cells (LECs), were marked by *Lcn2* expression (Figures 3C, D). Similar results were revealed from the CellMarker2.0 database query (33), where *Lcn2* was defined as a marker of murine lung neutrophils and AT2 cells (Supplementary Table S3). Importantly, the highest *Lcn2* expression was detected during the earlier inflammatory phase of the model (Figures 3E–G), which is characterized by epithelial damage and neutrophilic inflammation.

To validate the *in silico* mouse results, we examined *Lcn2* expression during the development of BLM-induced pulmonary inflammation and fibrosis. To this end, BLM (0.8 U/Kg) was administered by oropharyngeal aspiration to 8–10-wk-old C57Bl6 mice, which were then euthanized 3, 7, and 14 days post-BLM administration, timepoints corresponding to the inflammatory (3, 7) and fibrotic (14) phases of the disease (which resolves at 21 d; not shown). As expected, BLM administration resulted in the vascular leak and pulmonary edema, as indicated by the total protein concentration of the bronchoalveolar lavage fluid (BALF), determined with the Bradford assay (Figure 4A), as well as in inflammation, as indicated by the inflammatory cells in the (BALF) (Figure 4B). Soluble collagen levels in the BALF, as determined with the Direct Red assay, were also found gradually increasing in fibrotic lungs (Figure 1C). The H&E staining performed in lung sections of murine lungs post-BLM administration revealed the increasing presence of peribronchiolar and parenchymal fibrotic regions (Figure 4D). Moreover, the development of pulmonary fibrosis was reflected in the impairment of respiratory functions, as quantified with FlexiVent (Figures 4E–J). The development of BLM-induced pulmonary fibrosis and the impairment of respiratory functions were associated with increased lung tissue *Lcn2* mRNA expression, as detected with Q-RT-PCR, in all phases of the disease, but especially in the acute inflammatory phase (Figure 4K). A similar profile was detected in the *Lcn2* protein concentrations in the BALF (Figure 4L), while the increased concentration in the serum of the same mice was only detected in the acute phase. To possibly correlate *Lcn2* levels with immune cell populations in the BALF post-BLM administration, a multicolor FACS analysis was performed, quantifying 10 distinct immune cell types; the employed gating strategy is described in detail in Supplementary Figure S2. FACS indicated an abundance of neutrophils in the acute inflammatory phase post-BLM (Figure 4O), at the peak of *Lcn2* expression. However, increased *Lcn2* protein levels could still be detected in the fibrotic lung tissue 14 d post-BLM (Figures 4P, Q), while *Lcn2* immunostaining was localized in epithelial cells and fibrotic regions, constitutive expression was detected from the bronchial epithelium (Figure 4R).

To confirm *Lcn2* as a marker of pulmonary inflammation, we then examined *Lcn2* levels upon LPS-induced acute lung injury (ALI). LPS was administered (5 mL; 2 mg/mL) via a nebulizer (flowrate 4 lt/min) to WT C57Bl6 mice, that were euthanized 24 h later. The development of ALI, as indicated by the vascular leak (Figure 5A) and the infiltration of inflammatory cells (Figure 5B;

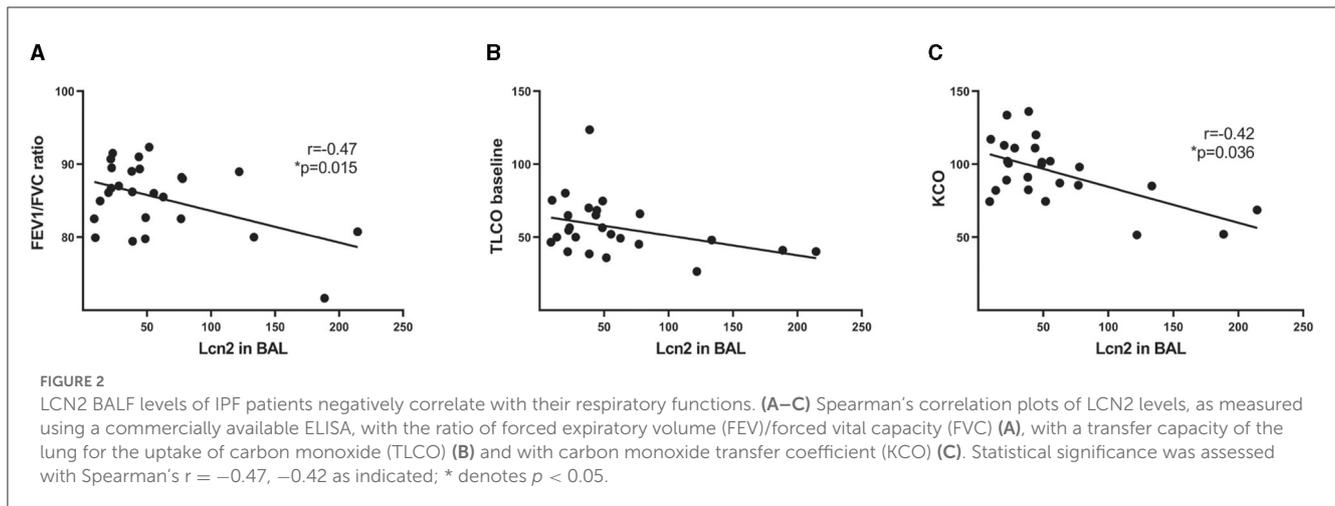


mostly neutrophils; data not shown) (35), was associated with increased *Lcn2* mRNA (Figure 5C) and protein levels (Figures 5D, E) in the lung tissue. The increased *Lcn2* expression upon ALI was also reflected in the BALF and sera of the same mice (Figures 5F, G).

Therefore, *Lcn2* is a marker of pulmonary inflammation in mice, correlating with epithelial damage and neutrophilic infiltration.

Genetic dissection of the role of *Lcn2* in pulmonary inflammation and fibrosis in mice

To dissect a possible role for *Lcn2* in pulmonary inflammation and fibrosis, we then investigated the effects of BLM-induced pulmonary inflammation and fibrosis on *Lcn2* ubiquitous knockout mice (KO); the lack of *Lcn2* expression in KO mice was verified



using Q-RT-PCR and Western blot analysis (Figures 6A, B). The BLM-induced weight loss, an indicator of overall systemic health, did not reach statistical significance in $Lcn2^{-/-}$ mice (Figure 6C); however, no statistically significant changes were detected in vascular leak (Figure 6D), inflammation (Figure 6E), or soluble BALF collagen (Figure 6F). Accordingly, no major statistically significant differences were detected in the associated distortion of lung architecture (Figures 6G, H). However, BLM-induced impairment of respiratory functions did not reach statistical significance in $Lcn2^{-/-}$ mice (Figures 6I, J), suggesting again, as the human data, a possible negative correlation of *Lcn2* expression with respiratory functions.

Given the suggested role of *Lcn2* in metabolic disorders and obesity (36) and the correlation between IPF and obesity in patients (37), the effect of obesity-driven microbiome changes in the lungs (38), as well as the suggested role of *Lcn2* in iron sequestration and microbiome regulation, we next investigated the role of *Lcn2* in the pathogenesis of pulmonary fibrosis in obese mice, following the high-fat diet (HFD) feeding for 13 weeks, in comparison with mice fed a matched control diet. No statistically significant changes in disease severity were observed either, although a clear trend of disease attenuation was observed (Supplementary Figures S3A–D), as opposed to lean mice.

Moreover, given the increased *Lcn2* expression in the acute phase post-BLM administration (Figure 4), as well as following LPS-induced ALI (Figure 5), we then examined a possible role of *Lcn2* in acute inflammation by administering LPS in $Lcn2^{-/-}$ and control wt mice. $Lcn2^{-/-}$ mice presented with increased pulmonary edema (Figure 7A), but no significant changes in inflammation (Figures 7B, C).

Therefore, despite observed trends in $Lcn2^{-/-}$ mice, no solid conclusions on the role of *Lcn2* in pulmonary inflammation and fibrosis could be drawn upon disease modeling in mice, in these settings.

Discussion

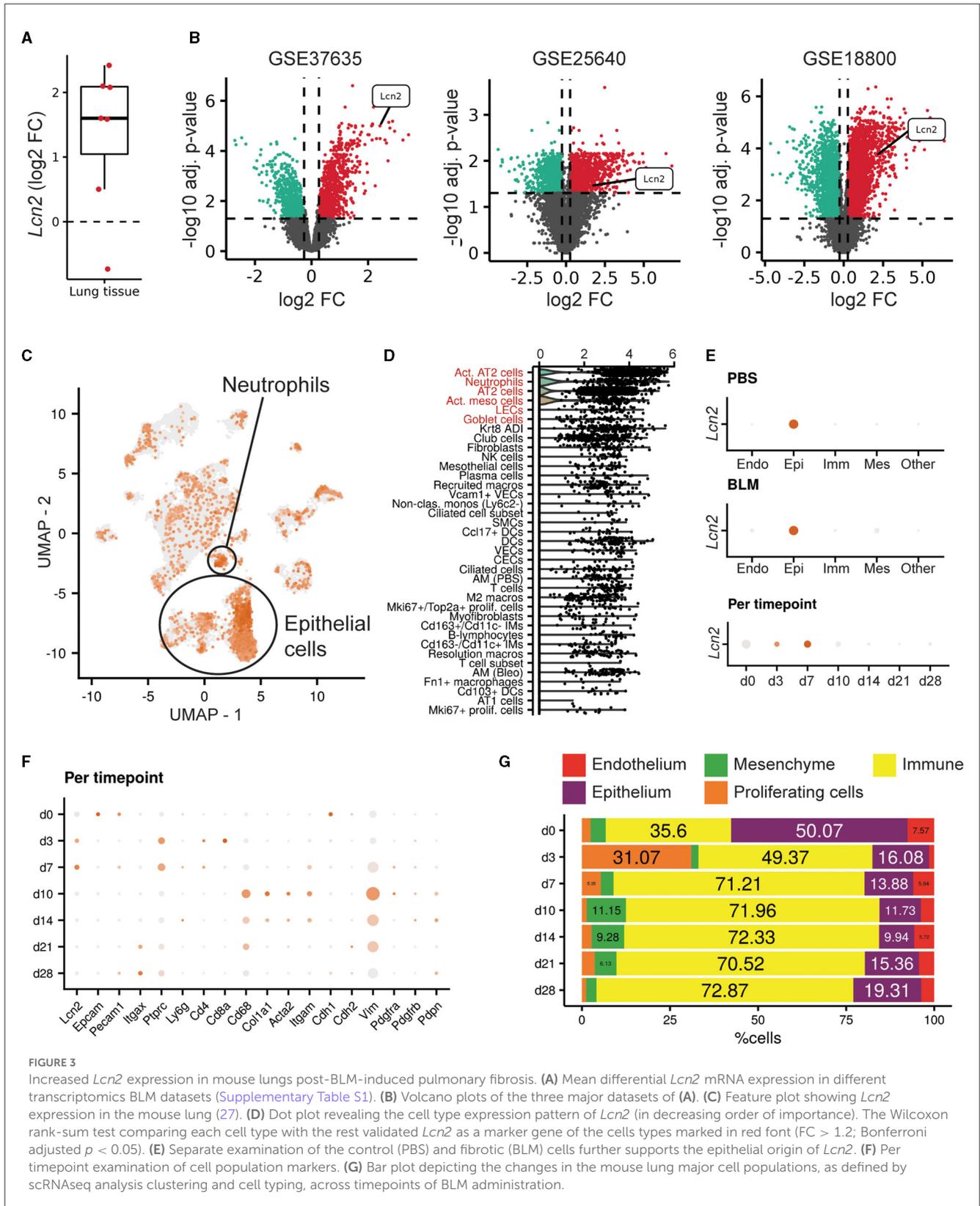
In this report, increased *LCN2* mRNA expression has been detected *in silico* in most available transcriptomics datasets at

Fibromine.com (Figure 1; Supplementary Table S1). The *in silico* approach, given the availability of datasets in Fibromine, emerges as a valuable surrogate tool for the identification of the expression levels of genes under investigation. Moreover, given the multiple available datasets/human samples, the method is more practical and valuable than the usual practice, i.e., individual RT-PCRs in a limited number of IPF samples.

LCN2 mRNA expression levels in IPF patients negatively correlated with respiratory functions (Figure 1); accordingly, *LCN2* BALF levels negatively correlated with patients' respiratory functions (FEV1/FVC, TLCO, and KCO) of a cohort ($n=26$) of IPF patients (Figure 2), in agreement with a previous study (17). However, much larger clinical studies will be needed to possibly associate *LCN2* expression levels, in both sera and BALF, with respiratory functions and other specific pathophysiological disease attributes. A meta-analysis of publicly available scRNAseq datasets indicated the lung epithelium as the major source of *LCN2* in the fibrotic lung (Figure 1), as previously shown with immunocytochemistry (17). Recurrent epithelial damage is considered the initiating insult of IPF pathogenesis, and acute exacerbation of IPF is characterized by increased alveolar epithelial cell injury, suggesting that future studies on *LCN2* and IPF should include the evaluation of *LCN2* levels in patients with acute IPF exacerbation and the correlation with other epithelial injury markers.

Similar results were obtained in BLM-induced pulmonary fibrosis in mice (Figures 3, 4), further indicating higher *Lcn2* expression in the acute phase of the disease, following BLM-induced epithelial damage and correlating with neutrophilic inflammation. *Lcn2* levels declined at the fibrotic phase, although remained higher than controls, as is the case for various inflammatory markers, e.g., TNF (39). Moreover, and in agreement with an acute role for *Lcn2*, neutrophilic infiltration upon LPS-induced ALI was also correlated with higher *Lcn2* expression (Figure 5).

However, despite the increased *Lcn2* expression upon BLM- or LPS-induced lung damage, no statistically significant changes were observed upon BLM or LPS administration to $Lcn2^{-/-}$ mice (Figures 6, 7; Supplementary Figure S3), suggesting that either *Lcn2* does not have a major pathogenic role or *Lcn2* can



have different roles in different cell populations, masked in the ubiquitous knockout mice, and that a cell-specific *Lcn2* deletion could be more informative. Moreover, it is also possible that a pathogenic role for *Lcn2* cannot be efficiently dissected in animal

models, as has been shown for many other genes (4). In this context, a very possible role of *Lcn2* in iron sequestration and microbiome regulation (18) cannot be likely examined in modeled mice, given their sterile and controlled living conditions, as well as due to the

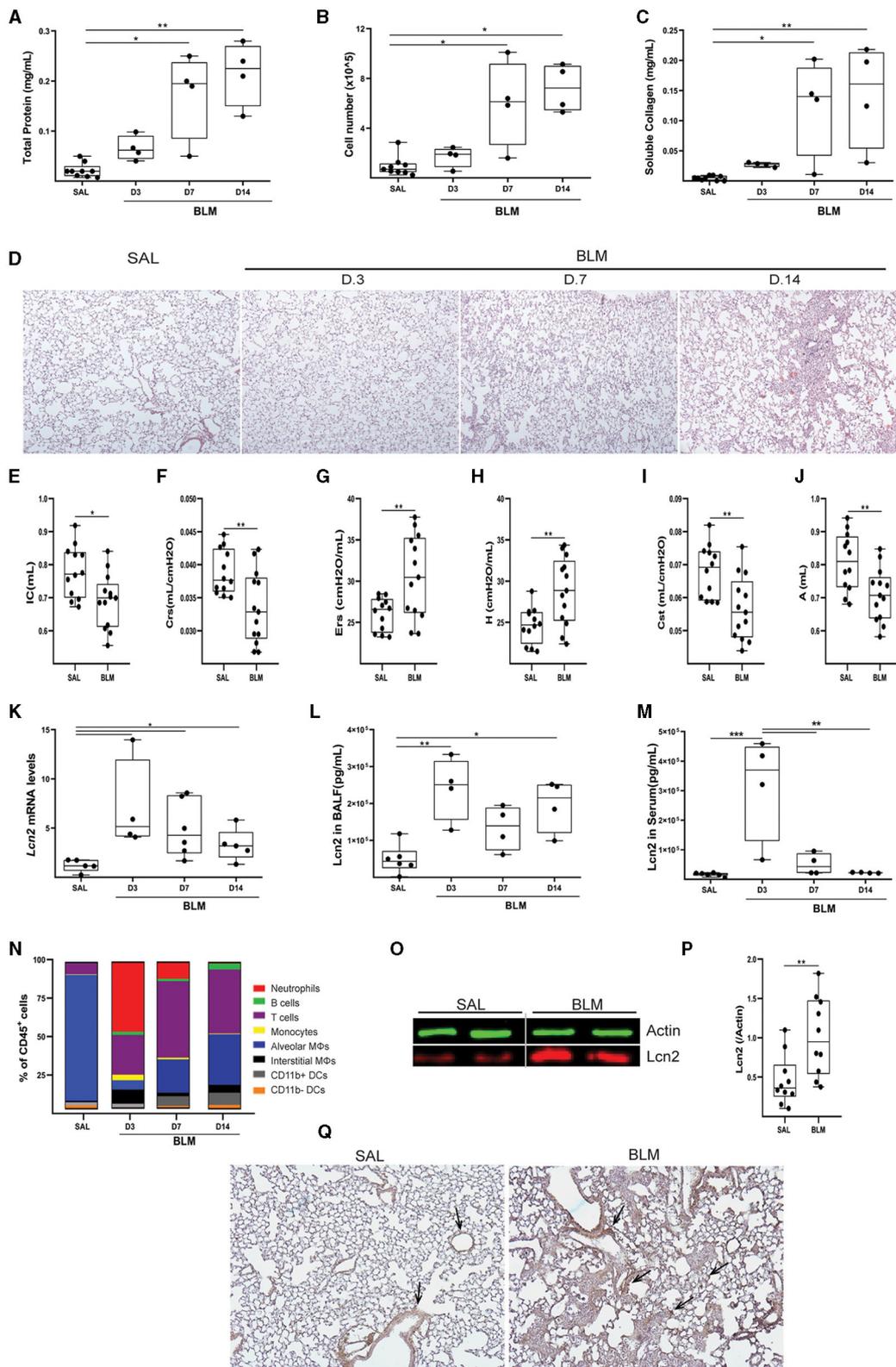


FIGURE 4 Increased *Lcn2* expression in mouse lungs during the development of BLM-induced pulmonary inflammation and fibrosis. **(A)** Total protein concentration in BALFs, as determined using the Bradford assay. **(B)** Inflammatory cell numbers in BALFs, as counted with a hemacytometer. **(C)** Soluble collagen levels in the BALFs as detected with the Direct Red assay. Statistical significance was assessed with one-way ANOVA; */** denote $p < 0.05/0.01$ respectively. **(D)** Representative images from H&E-stained lung sections of murine lungs at 3, 7, and 14 d post-BLM administration ($\times 10$). **(E–J)** Respiratory functions were measured with FlexiVent, 14 days post-BLM; mean respiratory system compliance (Crs); mean respiratory system elastance (Ers); mean tissue elastance (H); mean static lung compliance (Cst); mean total lung capacity (A). Cumulative results from three

(Continued)

FIGURE 4 (Continued)

independent experiments; statistical significance was assessed using the Mann–Whitney test; */** denote $p < 0.05/0.01$, respectively. (K) *Lcn2* mRNA expression was interrogated using Q-RT-PCR; Values were normalized over the expression of the housekeeping gene *B2m* and presented as fold change over control. (L, M) *Lcn2* concentration in BALF (L) and serum (M) of mice at 3, 7, and 14 d post-BLM administration. *Lcn2* levels were measured using a commercially available ELISA kit; Statistical significance was assessed with one-way ANOVA, */**/* denote $p < 0.05/0.01/0.001$ respectively. (N) Bar plot showing the percentage of immune cell populations in the murine lung post-BLM; the employed gating strategy is described in Supplementary Figure S2. (O) Representative Western blot of *Lcn2* expression (red) in fibrotic lungs, 14 d post-BLM. (P) Densitometry analysis of *Lcn2* expression, normalized to the expression of Actin (green); cumulative result from two independent experiments; statistical significance was assessed with unpaired t-test; ** denotes $p < 0.01$. (Q) Representative images of two independent experiments, from immunohistochemistry for *Lcn2* in control (SAL) and fibrotic (BLM) murine lung tissue ($\times 10$).

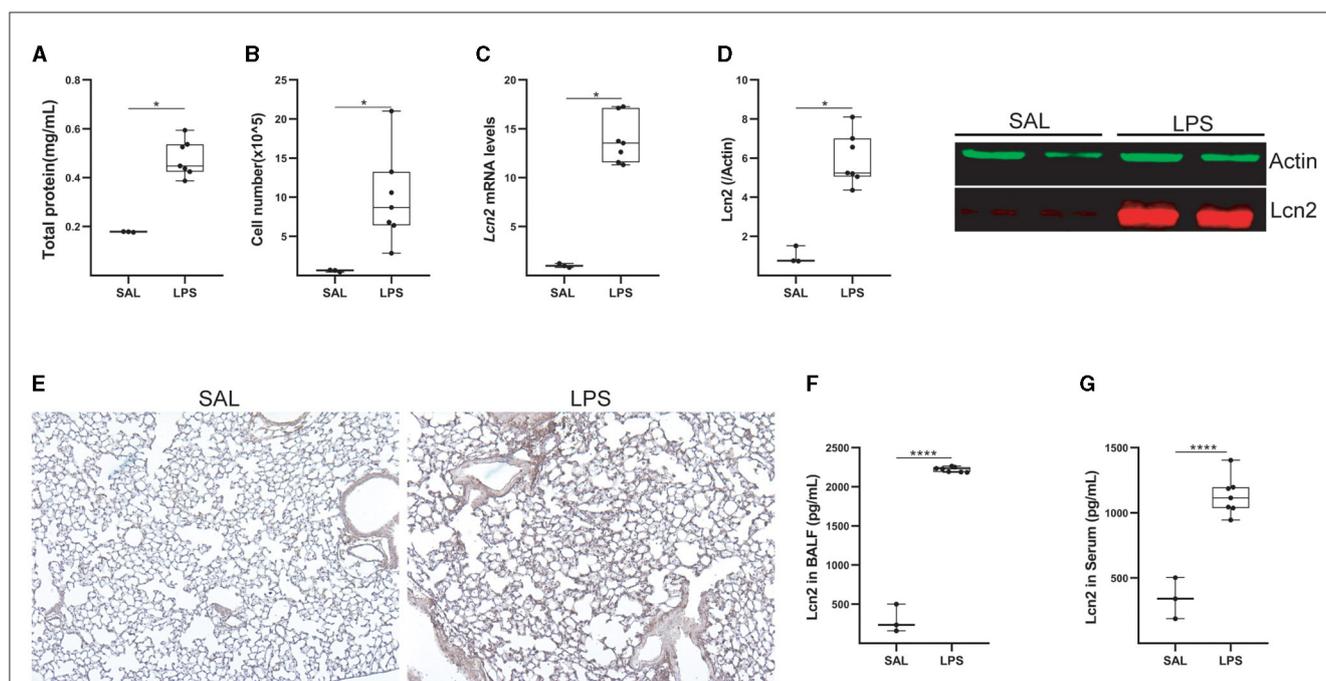


FIGURE 5

Lcn2 expression is upregulated during LPS-induced Acute Lung Injury (ALI). (A) Total protein concentration in BALF, as determined using the Bradford assay. (B) Inflammatory cell numbers in BALF, from saline and LPS-treated mice, as counted with a hemacytometer. (C) *Lcn2* mRNA expression was interrogated with Q-RT-PCR; Values were normalized over the expression of the housekeeping gene *B2m* and presented as fold change over control; representative results from three independent experiments. (D) Western blot of *Lcn2* expression (red) in lungs from mice with LPS-induced ALI, followed by densitometry analysis of *Lcn2* expression, normalized to the expression of Actin (green). (E) Representative images from immunohistochemistry for *Lcn2* in lungs from control (SAL) and LPS-treated mice ($\times 10$). (F, G) *Lcn2* levels in BALF (F) and serum (G) of mice were estimated using ELISA; statistical significance was assessed using the Mann–Whitney test; */**** denote $p < 0.05/0.0001$.

species populating the lung that are not amenable to the suggested bacteriostatic functions of *Lcn2* (38, 40). However, a role for *LCN2* in microbiome regulation in humans remains likely and should be pursued in future clinical studies, especially since increased airway microbiota has been associated with a more rapid disease progression and a higher risk of mortality across different patient cohorts and quantification platforms (20, 41, 42).

Moreover, microbiome differences in different animal houses could explain the contradictory results on the role of *Lcn2* in inflammation in mice. For example, *Lcn2* has been suggested to mediate the recruitment of neutrophils and thus to stimulate pro-inflammatory signaling; however, anti-inflammatory effects have also been suggested, including *M2* polarization and T_{Reg} s expansion (10). We reported here no major role for *Lcn2* in LPS-induced ALI, while it was recently reported that *Lcn2*^{-/-} mice had relatively increased survival than control mice following intratracheal administration of LPS (43); the contradiction could

be due to experimental design, dose, and species of administered LPS, as well as the local microbiome of the animal houses. In the same context, systemic administration of LPS in *Lcn2*^{-/-} mice was reported to result in exacerbated neuroinflammatory responses (44), although an opposite role in neuroinflammation has been also suggested promoting macrophage *M1* polarization (45). However, in the lungs, *LCN2* was reported to deactivate macrophages resulting in impaired immune responses following pneumococcal pneumonia (46).

As an acute phase response protein, secreted by epithelial cells upon damage, and/or infiltrating neutrophils, it is conceivable that *LCN2* may contribute to chronic damage responses via the lung epithelium in IPF patients through the amplification of neutrophil recruitment. Increased neutrophils were detected in the IPF cohort examined here (Table 1; $p = 0.038$), while BAL neutrophilia has been proposed as an independent predictor of early mortality in IPF patients (47).

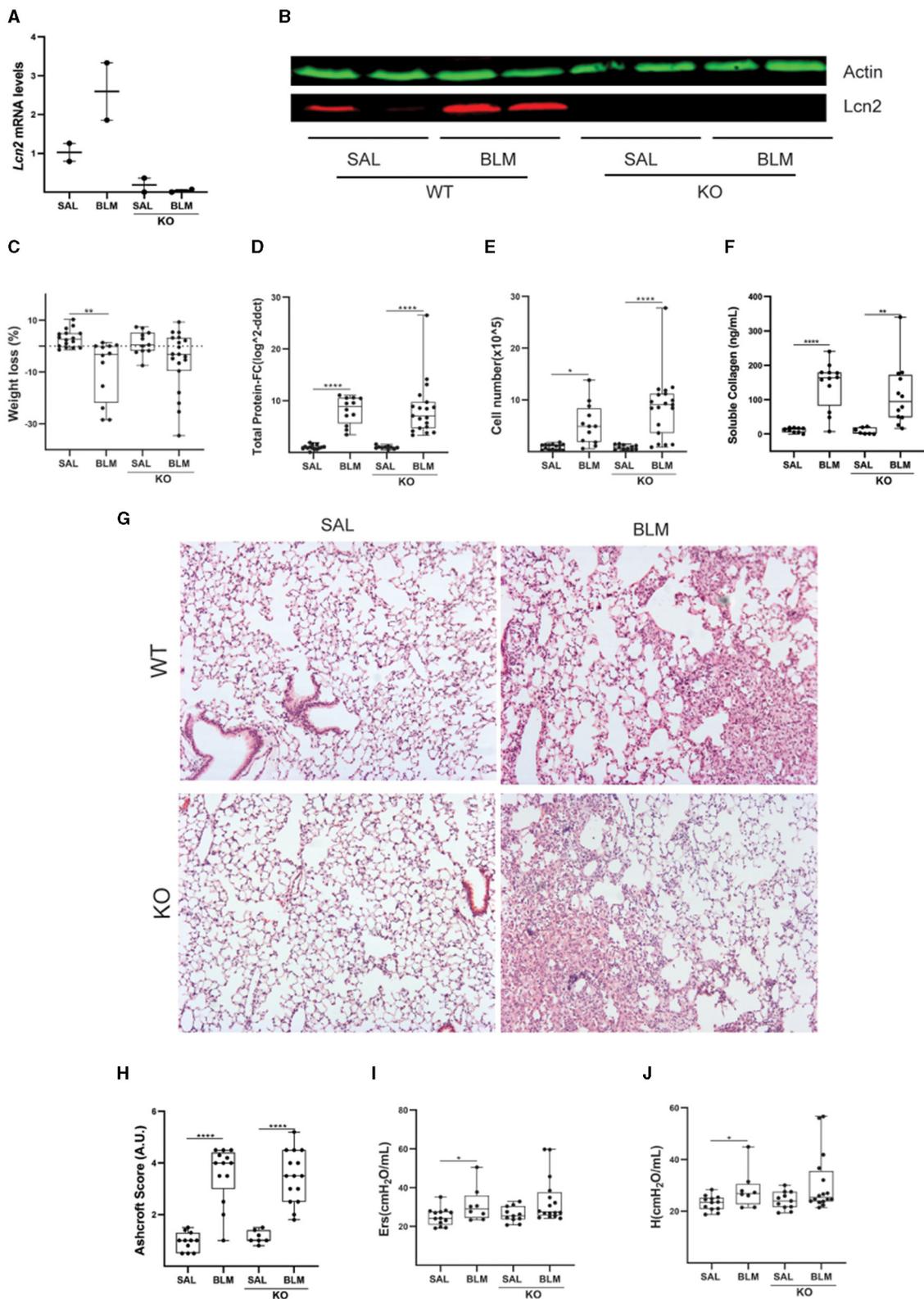
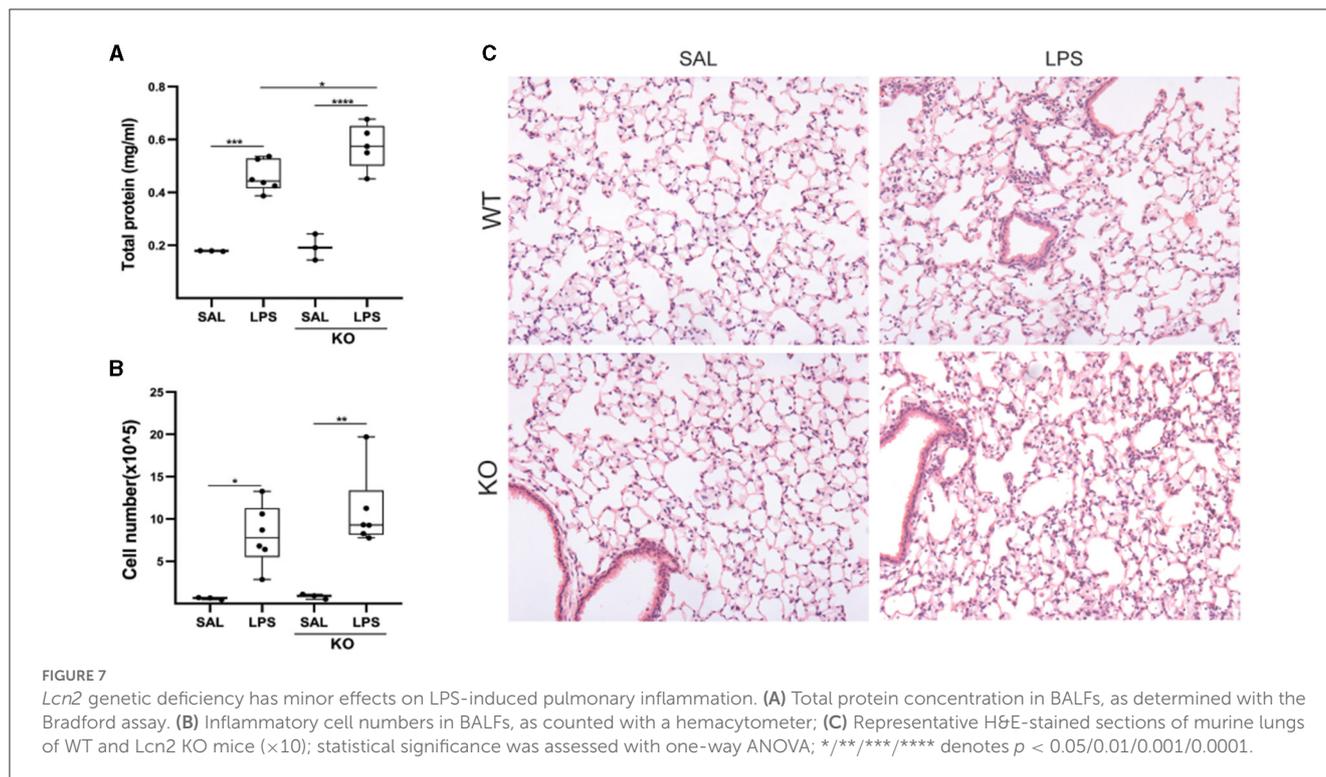


FIGURE 6 *Lcn2* genetic deficiency has minor effects in bleomycin (BLM)-induced pulmonary inflammation and fibrosis. **(A)** *Lcn2* mRNA expression was interrogated with Q-RT-PCR; Values were normalized over the expression of the housekeeping gene *B2m* and presented as fold change over control; **(B)** Representative Western blot of *Lcn2* expression (red) in lungs from WT and KO mice treated with BLM confirming the global depletion of *Lcn2* in KO mice. **(C)** Weight loss post-BLM administration. **(D)** Total protein concentration in BALFs, as determined with the Bradford assay. **(E)** Inflammatory cell numbers in BALFs, as counted with a hemacytometer. **(F)** Soluble collagen levels in the BALFs as detected with the Direct Red assay. **(G)** Representative H&E-stained lung sections ($\times 10$). **(H)** Ashcroft scoring of disease severity. **(I, J)** Indicated respiratory functions were measured with FlexiVent; statistical significance was assessed with one-way ANOVA; */**/**** denotes $p < 0.05/0.01/0.0001$.



Additionally, a high neutrophil to lymphocyte ratio (NLR) as measured from complete blood counts has also been associated with increased mortality in IPF (48). LCN2 has been shown to promote the formation of neutrophil extracellular traps (NETs) (49), which have been implicated in the pathogenesis of several diseases including IPF (50). In skin psoriasis, the amplification loop of LCN2 parallel to neutrophil-produced extracellular NETs was shown to participate in the enhancement and persistence of the local inflammatory response (51). The proinflammatory activity of NETs and LCN2 induction in psoriasis was suggested to be dependent on TLR4/IL-36R crosstalk and MyD88/nuclear factor-kappa B (NF- κ B) downstream signaling (51).

Overall, although the possible role for LCN2 in IPF pathogenesis remains obscure, the acute increase in *Lcn2* expression following both LPS-induced ALI and BLM-induced pulmonary inflammation and fibrosis suggests that *Lcn2* is an acute phase protein of lung damage in mice, as previously suggested for acute kidney injury (9) and acute exacerbation of cystic fibrosis (52), correlating with epithelial damage and neutrophilic infiltration. Moreover, the increased LCN2 mRNA levels detected in IPF patients suggest that LCN2 levels can be used as surrogate biomarkers of pulmonary inflammation and a possible indicator of compromised pulmonary functions, urging for larger studies.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics Committees of the University Hospital of Heraklion (IRB numbers: 1045 and 17030). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin. The animal study was approved by Institutional Animal Ethical Committee (IAEC) of Biomedical Sciences Research Center Alexander Fleming (#373/375), as well as the Veterinary Service and Fishery Department of the local governmental prefecture (#5508).

Author contributions

AG performed most presented experiments and analyzed the relative data, assisted by IB, PK, KT, TK, and AT. KN and SG performed FACS. ET performed human ELISAs. DF performed *in silico* data re-analysis and supervised all statistical analyses. KA led the relative clinical protocol and provided all human samples. AG, DE, ET, and VA wrote the article, which was critically read by all co-authors. All authors contributed to the article and approved the submitted version.

Funding

AG was supported by a fellowship from the Hellenic Foundation for Research and Innovation (#789). The research

was further partly supported through the Hellenic Foundation for Research and Innovation (HFRI) under the 2nd Call for HFRI Research Projects to support Faculty Members and Researchers (#3565 to VA). The funders had no role in the design of the study, collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Martinez FJ, Collard HR, Pardo A, Raghu G, Richeldi L, Selman M, et al. Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers*. (2017) 3:17074. doi: 10.1038/nrdp.2017.74
- Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, et al. Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med*. (2018) 198:e44–68. doi: 10.1164/rccm.201807-1255ST
- Vukmirovic M, Kaminski N. Impact of transcriptomics on our understanding of pulmonary fibrosis. *Front Med (Lausanne)*. (2018) 5:87. doi: 10.3389/fmed.2018.00087
- Mouratis MA, Aidinis V. Modeling pulmonary fibrosis with bleomycin. *Curr Opin Pulm Med*. (2011) 17:355–61. doi: 10.1097/MCP.0b013e328349ac2b
- Fanidis D, Moulos P, Aidinis V. Fibromine is a multi-omics database and mining tool for target discovery in pulmonary fibrosis. *Sci Rep*. (2021) 11:21712. doi: 10.1038/s41598-021-01069-w
- Fanidis D, Pezoulas VC, Fotiadis D, Aidinis V. An explainable machine learning-driven proposal of pulmonary fibrosis biomarkers. *Comput Struct Biotechnol J*. (2023) 21:2305–15. doi: 10.1016/j.csbj.2023.03.043
- Kjeldsen L, Johnsen AH, Sengelov H, Borregaard N. Isolation primary structure of NGAL, a novel protein associated with human neutrophil gelatinase. *J Biol Chem*. (1993) 268:10425–32. doi: 10.1016/S0021-9258(18)82217-7
- Flower DR, North AC, Sansom EC. The lipocalin protein family: structural and sequence overview. *Biochim Biophys Acta*. (2000) 1482:9–24. doi: 10.1016/S0167-4838(00)00148-5
- Devarajan P. Neutrophil gelatinase-associated lipocalin: a promising biomarker for human acute kidney injury. *Biomark Med*. (2010) 4:265–80. doi: 10.2217/bmm.10.12
- Guardado S, Ojeda-Juárez D, Kaul M, Nordgren MT. Comprehensive review of lipocalin 2-mediated effects in lung inflammation. *Am J Physiol-Lung Cell Mol Physiol*. (2021) 321:L726–33. doi: 10.1152/ajplung.00080.2021
- Betsuyaku T, Nishimura M, Takeyabu K, Tanino M, Venge P, Xu S, et al. Neutrophil granule proteins in bronchoalveolar lavage fluid from subjects with subclinical emphysema. *Am J Respir Crit Care Med*. (1999) 159:1985–91. doi: 10.1164/ajrccm.159.6.9809043
- Eagan TM, Damás JK, Ueland T, Voll-Aanerud M, Mollnes TE, Hardie JA, et al. Neutrophil gelatinase-associated lipocalin: a biomarker in COPD. *Chest*. (2010) 138:888–95. doi: 10.1378/chest.09-2718
- Treekitkarnmongkol W, Hassane M, Sinjab A, Chang K, Hara K, Rahal Z, et al. Augmented lipocalin-2 is associated with COPD and counteracts lung adenocarcinoma development. *Am J Respir Crit Care Med*. (2020). doi: 10.1164/rccm.202004-1079OC
- Kangelaris KN, Prakash A, Liu KD, Aouizerat B, Woodruff PG, Erle DJ, et al. Increased expression of neutrophil-related genes in patients with early sepsis-induced ARDS. *Am J Physiol Lung Cell Mol Physiol*. (2015) 308:L1102–1113. doi: 10.1152/ajplung.00380.2014
- Huang Z, Li H, Liu S, Jia J, Zheng Y, Cao B. Identification of neutrophil-related factor LCN2 for predicting severity of patients with influenza A virus and SARS-CoV-2 infection. *Front Microbiol*. (2022) 13:854172. doi: 10.3389/fmicb.2022.854172
- Ntatsoulis K, Karamitsakos T, Tsitoura E, Stylianaki EA, Matralis AN, Tzouvelekas A, et al. Commonalities between ARDS, pulmonary fibrosis and COVID-19: the potential of autotaxin as a therapeutic target. *Front Immunol*. (2021) 12:687397. doi: 10.3389/fimmu.2021.687397
- Ikezoe K, Handa T, Mori K, Watanabe K, Tanizawa K, Aihara K, et al. Neutrophil gelatinase-associated lipocalin in idiopathic pulmonary fibrosis. *Eur Respir J*. (2014) 43:1807–9. doi: 10.1183/09031936.00192613
- Flo TH, Smith KD, Sato S, Rodriguez DJ, Holmes MA, Strong RK, et al. Lipocalin 2 mediates an innate immune response to bacterial infection by sequestering iron. *Nature*. (2004) 432:917–21. doi: 10.1038/nature03104
- Ali MK, Kim RY, Brown AC, Donovan C, Vanka KS, Mayall JR, et al. Critical role for iron accumulation in the pathogenesis of fibrotic lung disease. *J Pathol*. (2020). doi: 10.1002/path.5401
- Molyneaux PL, Cox MJ, Willis-Owen SA, Mallia P, Russell KE, Russell AM, et al. The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. (2014) 190:906–13. doi: 10.1164/rccm.201403-0541OC
- Lynch DA, Sverzellati N, Travis WD, Brown KK, Colby TV, Galvin JR, et al. Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper. *Lancet Respir Med*. (2018) 6:138–53. doi: 10.1016/S2213-2600(17)30433-2
- Barbayanni I, Ninou I, Tzouvelekas A, Aidinis V. Bleomycin revisited: a direct comparison of the intratracheal micro-spraying and the oropharyngeal aspiration routes of bleomycin administration in mice. *Front Med (Lausanne)*. (2018) 5:269. doi: 10.3389/fmed.2018.00269
- Mouratis MA, Magkrioti C, Oikonomou N, Katsifa A, Prestwich GD, Kaffe E, et al. Autotaxin and Endotoxin-Induced Acute Lung Injury. *PLoS One*. (2015) 10:e0133619. doi: 10.1371/journal.pone.0133619
- Adams TS, Schupp JC, Poli S, Ayaub EA, Neumark N, Ahangari F, et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv*. (2020) 6:eaba1983. doi: 10.1126/sciadv.aba1983
- Habermann AC, Gutierrez AJ, Bui LT, Yahn SL, Winters NI, Calvi CL, et al. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci Adv*. (2020) 6: aba1972. doi: 10.1126/sciadv.aba1972
- Mayr CH, Simon LM, Leuschner G, Ansari M, Schniering J, Geyer PE, et al. Integrative analysis of cell state changes in lung fibrosis with peripheral protein biomarkers. *EMBO Mol Med*. (2021) 13:e12871. doi: 10.15252/emmm.202012871
- Strunz M, Simon LM, Ansari M, Kathiriya JJ, Angelidis I, Mayr CH, et al. Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nat Commun*. (2020) 11:3559. doi: 10.1038/s41467-020-17358-3
- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive integration of single-cell data. *Cell*. (2019) 177:1888–902.e1821. doi: 10.1016/j.cell.2019.05.031
- Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell*. (2021) 184:3573–87.e3529. doi: 10.1016/j.cell.2021.04.048
- Zhuang H, Wang H, Ji Z. findPC: An R package to automatically select the number of principal components in single-cell analysis. *Bioinformatics*. (2022) 38:2949–51. doi: 10.1093/bioinformatics/btac235

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2023.1195501/full#supplementary-material>

31. McInnes L, John H, James M. UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv*. (2020) 1802.03426. doi: 10.48550/arXiv.1802.03426
32. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. (2019) 37:38–44. doi: 10.1038/nbt.4314
33. Hu C, Li T, Xu Y, Zhang X, Li F, Bai J, et al. CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res*. (2023) 51:D870–D876. doi: 10.1093/nar/gkac947
34. Tashiro J, Rubio GA, Limper AH, Williams K, Elliot SJ, Ninou I, et al. Exploring animal models that resemble idiopathic pulmonary fibrosis. *Front Med (Lausanne)*. (2017) 4:118. doi: 10.3389/fmed.2017.00118
35. Chen H, Bai C, Wang X. The value of the lipopolysaccharide-induced acute lung injury model in respiratory medicine. *Expert Rev Respir Med*. (2010) 4:773–83. doi: 10.1586/ers.10.71
36. Jaber SA, Cohen A, D'Souza C, Abdulrazzaq YM, Ojha S, Bastaki S, et al. Lipocalin-2: structure, function, distribution and role in metabolic disorders. *Biomed Pharmacother*. (2021) 142:112002. doi: 10.1016/j.biopha.2021.112002
37. Guo X, Sunil C, Qian G. Obesity the development of lung fibrosis. *Front Pharmacol*. (2021) 12:812166. doi: 10.3389/fphar.2021.812166
38. Galaris A, Fanidis D, Stylianaki EA, Harokopos V, Kalantzi AS, Moulos P, et al. Obesity reshapes the microbial population structure along the gut-liver-lung axis in mice. *Biomedicines*. (2022) 10:2. doi: 10.3390/biomedicines10020494
39. Oikonomou N, Harokopos V, Zalevsky J, Valavanis C, Kotanidou A, Szymkowski DE, et al. Soluble TNF mediates the transition from pulmonary inflammation to fibrosis. *PLoS ONE*. (2006) 1:e108. doi: 10.1371/journal.pone.0000108
40. Xiao X, Yeoh BS, Vijay-Kumar M. Lipocalin 2: an emerging player in iron homeostasis and inflammation. *Annu Rev Nutr*. (2017) 37:103–30. doi: 10.1146/annurev-nutr-071816-064559
41. Han MK, Zhou Y, Murray S, Tayob N, Noth I, Lama VN, et al. Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. *Lancet Respir Med*. (2014) 2:548–56. doi: 10.1016/S2213-2600(14)70069-4
42. O'Dwyer DN, Ashley SL, Gurczynski SJ, Xia M, Wilke C, Falkowski NR, et al. Lung microbiota contribute to pulmonary inflammation and disease progression in pulmonary fibrosis. *Am J Respir Crit Care Med*. (2019) 199:1127–38. doi: 10.1164/rccm.201809-1650OC
43. An HS, Yoo JW, Jeong JH, Heo M, Hwang SH, Jang HM, et al. Lipocalin-2 promotes acute lung inflammation and oxidative stress by enhancing macrophage iron accumulation. *Int J Biol Sci*. (2023) 19:1163–77. doi: 10.7150/ijbs.79915
44. Kang SS, Ren Y, Liu CC, Kurti A, Baker KE, Bu G, et al. Lipocalin-2 protects the brain during inflammatory conditions. *Mol Psychiatry*. (2018) 23:344–50. doi: 10.1038/mp.2016.243
45. Jang E, Lee S, Kim JH, Kim JH, Seo JW, Lee WH, et al. Secreted protein lipocalin-2 promotes microglial M1 polarization. *FASEB J*. (2013) 27:1176–90. doi: 10.1096/fj.12-222257
46. Warszawska JM, Gawish R, Sharif O, Sigel S, Doninger B, Lakovits K, et al. Lipocalin 2 deactivates macrophages and worsens pneumococcal pneumonia outcomes. *J Clin Invest*. (2013) 123:3363–72. doi: 10.1172/JCI67911
47. Kinder BW, Brown KK, Schwarz MI, Ix JH, Kervitsky A, King TE. Baseline BAL neutrophilia predicts early mortality in idiopathic pulmonary fibrosis. *Chest*. (2008) 133:226–32. doi: 10.1378/chest.07-1948
48. Mikolasch TA, George PM, Sahota J, Nancarrow T, Barratt SL, Woodhead FA, et al. Multi-center evaluation of baseline neutrophil-to-lymphocyte (NLR) ratio as an independent predictor of mortality and clinical risk stratifier in idiopathic pulmonary fibrosis. *EClinicalMedicine*. (2023) 55:101758. doi: 10.1016/j.eclinm.2022.101758
49. Li H, Feng D, Cai Y, Liu Y, Xu M, Xiang X, et al. Hepatocytes and neutrophils cooperatively suppress bacterial infection by differentially regulating lipocalin-2 and neutrophil extracellular traps. *Hepatology*. (2018) 68:1604–20. doi: 10.1002/hep.29919
50. Khawaja AA, Chong DLW, Sahota J, Mikolasch TA, Pericleous C, Ripoll VM, et al. Identification of a novel HIF-1 α (M) β (2) integrin-NET axis in fibrotic interstitial lung disease. *Front Immunol*. (2020) 11:2190. doi: 10.3389/fimmu.2020.02190
51. Shao S, Fang H, Dang E, Xue K, Zhang J, Li B, et al. Neutrophil extracellular traps promote inflammatory responses in psoriasis via activating epidermal TLR4/IL-36R crosstalk. *Front Immunol*. (2019) 10:746. doi: 10.3389/fimmu.2019.00746
52. Eichler I, Nilsson M, Rath R, Enander I, Venge P, Koller YD. Human neutrophil lipocalin, a highly specific marker for acute exacerbation in cystic fibrosis. *Eur Respir J*. (1999) 14:1145–9. doi: 10.1183/09031936.99.14511459

SRC and TKS5 mediated podosome formation in fibroblasts promotes extracellular matrix invasion and pulmonary fibrosis

Η διαμεσολάβηση των SRC και TKS5 στον σχηματισμό ποδοσωμάτων από ινοβλάστες προωθεί εισβολή στην εξωκυτταρική μήτρα και πνευμονική ίνωση

Η ενεργοποίηση και συσσώρευση πνευμονικών οινόβλαστών έχει ως αποτέλεσμα την ασυνήθιστη συσσώρευση στοιχείων της εξωκυτταρικής μήτρας, ένα παθολογικό χαρακτηριστικό της ιδιοπαθούς πνευμονικής ίνωσης, μιας θανάσιμης και ανίατης ασθένειας. Σε αυτήν την αναφορά, αυξημένη έκφραση της TKS5, μιας ικριωματικής πρωτεΐνης απαραίτητης για τον σχηματισμό ποδοσωμάτων, ανιχνεύθηκε στον πνευμονικό ιστό ασθενών ιδιοπαθούς πνευμονικής ίνωσης (IPF) και ποντικών που είχαν λάβει μπλεομυκίνη (BLM). Το προϊνωτικό περιβάλλον βρέθηκε να προάγει την έκφραση του TKS5 και τον σχηματισμό προεξεχόντων ροζετών ποδοσωμάτων από πνευμονικούς ινοβλάστες. Οι ροζέτες αυτές διατηρήθηκαν *ex vivo* καταλήγοντας σε αυξημένη εισβολή στην εξωκυτταρική μήτρα. Tks5^{+/-} ποντικοί βρέθηκαν ανθεκτικοί στην πρόκληση πνευμονικής ίνωσης από BLM, γεγονός που μπορεί εν πολλοίς να αποδοθεί σε εξασθενημένο σχηματισμό ινοβλαστικών ποδοσωμάτων και σε μειωμένη εισβολή τους στην εξωκυτταρική μήτρα. Όπως προβλέφθηκε υπολογιστικά, δείχθηκε πως αναστολή της SRC κινάσης είναι ικανή να περιορίσει τον σχηματισμό ποδοσωμάτων από πνευμονικούς ινοβλάστες, την εισβολή της εξωκυτταρικής μήτρας αλλά και την πνευμονική ίνωση που προκαλείται από BLM, προτείνοντας την φαρμακολογική στόχευση των ποδοσωμάτων ως μία πολλά υποσχόμενη θεραπευτική επιλογή για την πνευμονική ίνωση.

SRC and TKS5 mediated podosome formation in fibroblasts promotes extracellular matrix invasion and pulmonary fibrosis

Received: 5 February 2023

Accepted: 11 September 2023

Published online: 21 September 2023

 Check for updates

Ilianna Barbayianni^{1,7}, Paraskevi Kanellopoulou^{1,7}, Dionysios Fanidis¹, Dimitris Nastos¹, Eleftheria-Dimitra Ntouskou¹, Apostolos Galaris¹, Vaggelis Harokopos¹, Pantelis Hatzis¹, Eliza Tsitoura², Robert Homer³, Naftali Kaminski⁴, Katerina M. Antoniou², Bruno Crestani⁵, Argyrios Tzouvelelis⁶ & Vassilis Aidinis¹✉

The activation and accumulation of lung fibroblasts resulting in aberrant deposition of extracellular matrix components, is a pathogenic hallmark of Idiopathic Pulmonary Fibrosis, a lethal and incurable disease. In this report, increased expression of TKS5, a scaffold protein essential for the formation of podosomes, was detected in the lung tissue of Idiopathic Pulmonary Fibrosis patients and bleomycin-treated mice. The profibrotic milieu is found to induce *TKS5* expression and the formation of prominent podosome rosettes in lung fibroblasts, that are retained ex vivo, culminating in increased extracellular matrix invasion. *Tks5*^{-/-} mice are found resistant to bleomycin-induced pulmonary fibrosis, largely attributed to diminished podosome formation in fibroblasts and decreased extracellular matrix invasion. As computationally predicted, inhibition of src kinase is shown to potently attenuate podosome formation in lung fibroblasts and extracellular matrix invasion, and bleomycin-induced pulmonary fibrosis, suggesting pharmacological targeting of podosomes as a very promising therapeutic option in pulmonary fibrosis.

Tissue fibrosis is a pathogenic process that affects most organs and constitutes a complication of many chronic diseases including cancer; such fibroproliferative disorders account for >45% of all disease-related deaths worldwide¹. Among them, Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive, interstitial lung disease affecting mostly older adults. IPF patients exhibit progressive worsening of respiratory functions, which lead to dyspnea and eventually to respiratory failure.

Histologically, IPF is characterized by lung parenchymal scarring, as evident by a usual interstitial pneumonia (UIP) profile, characterized by patchy dense fibrosis with architectural distortion and a subpleural and paraseptal preference, and is distinguished by the presence of fibroblast foci². Although the etiopathogenesis of IPF remains largely elusive, the prevailing hypothesis suggests that the mechanisms driving IPF involve age-related aberrant recapitulation of developmental

¹Institute for Fundamental Biomedical Research, Biomedical Sciences Research Center Alexander Fleming, Athens, Greece. ²Department of Respiratory Medicine, School of Medicine, University of Crete, Heraklion, Greece. ³Department of Pathology, Yale School of Medicine, New Haven, CT, USA. ⁴Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA. ⁵Department of Pulmonology, Bichat-Claude Bernard Hospital, Paris, France. ⁶Department of Respiratory Medicine, School of Medicine, University of Patras, Patras, Greece. ⁷These authors contributed equally: Ilianna Barbayianni, Paraskevi Kanellopoulou. ✉ e-mail: V.Aidinis@Fleming.gr

programs and reflect abnormal, deregulated wound healing in response to persistent alveolar epithelial damage, resulting in the accumulation of lung fibroblasts (LFs)³.

LFs are the main effector cells in pulmonary fibrosis, secreting exuberant amounts of extracellular matrix (ECM) components, such as different types of collagens. LFs also secrete a variety of ECM remodeling enzymes, such as matrix metalloproteinases (MMPs), thus coordinating the overall ECM structural organization and consequently the mechanical properties of the lung⁴. LF activation upon fibrogenic cues, such as TGF β or other growth factors (including PDGF and VEGF), is characterized by the expression of alpha smooth muscle actin (α SMA/ACTA2), and/or increased collagen expression, as exemplified by COL1A1^{5,6}. ECM fibrotic remodeling and resulting mechanical cues are considered as crucial stimulating and perpetuating factors for LF activation^{6,7}, while the chemoattraction of LFs to various signals and their resistance to apoptosis has been suggested to promote respectively their recruitment and accumulation^{5,6}.

Fibroblast accumulation in pulmonary fibrosis has also been suggested to be mediated by their ability to invade the underlying ECM, and increased ECM invasion of fibroblasts isolated from the lung tissue of IPF patients or animal models has been reported^{8–11}. Activation of invasion, critical for embryonic development, is among the well-established hallmarks of cancer¹², and one of the many shared hallmarks between cancer cells and activated LFs¹³. Invasion critically relies on the proteolysis of the underlying ECM via invadopodia in cancer cells and podosomes in other cell types^{14,15}.

Podosomes are comprised of a filamentous (F)-actin-rich core enriched in actin-regulating proteins, such as the Arp2/3 complex and cortactin (CTTN), and are surrounded by a ring of scaffold proteins, most notably SH3 and PX domains 2A (SH3PXD2A; commonly known as tyrosine kinase substrate with 5 SH3 domains, TKS5)^{14–16}. The effector molecules of the podosomes are various proteases, such as matrix metalloproteinases (MMPs, mainly 2, 9, and 14) that digest the ECM locally, thus stimulating the invasion and migration of podosome bearing cells^{16,17}.

Tks5 expression is necessary for neural crest cell migration during embryonic development in zebrafish¹⁸, and homozygous disruption of *Tks5* in mice resulted in neonatal death¹⁹. Beyond embryonic development, which heavily relies on migration and invasion, increased TKS5 expression has been reported in different types of cancers^{14–16}, including lung adenocarcinoma, where it was suggested to mediate metastatic invasion²⁰. Pulmonary fibrosis confers one of the highest risks for lung cancer development, while many similarities between activated LFs and cancer cells have been suggested, including ECM invasion¹³. Therefore, in this report we investigated a possible role of TKS5 and podosomes in the pathogenesis of pulmonary fibrosis employing in silico analysis of publicly available human and mouse transcriptomic datasets, de novo analysis of human samples and associated clinical data, disease modeling in mice, ex vivo/in vitro human/mouse cell cultures and dedicated functional assays, as well as pharmacologic validation experiments. In this context, increased *Tks5* expression is detected in both human and mouse fibrotic lungs, primarily expressed in LFs. The profibrotic milieu, is shown to induce TKS5 expression and the formation of prominent podosome rosettes in fibroblasts, culminating in increased ECM invasion. Haploinsufficient *Tks5*^{+/−} mice are found resistant to BLM-induced pulmonary fibrosis, largely attributable to diminished podosome formation in LFs and decreased ECM invasion. Expression profiling reveals an ECM-podosome cross talk, and pharmacologic connectivity map analysis suggests several inhibitors that could prevent podosome formation and thus pulmonary fibrosis. Among them, inhibition of src kinase is shown to potently attenuate podosome formation in LFs, ECM invasion, as well as BLM-induced pulmonary fibrosis.

Results

Increased *TKS5* expression in pulmonary fibrosis

Increased *TKS5* mRNA levels were detected in silico in the lung tissue of IPF patients as compared with control samples (Fig. 1a), in most publicly available IPF transcriptomic datasets (Supplementary Table 1) at Fibromine²¹, including three of the largest ones (Fig. 1b and Supplementary Fig. 1a, c). Importantly, *TKS5* mRNA expression in fibrotic lungs correlated with the expression of *COL1A1* (Fig. 1c and Supplementary Fig. 1b, d), a well-established marker of fibrotic gene expression. Confirming the in silico results, increased *TKS5* mRNA levels were detected with quantitative RNA RT-PCR (Q-RT-PCR) in lung tissue isolated from IPF patients ($n=20$), as compared with COPD patients ($n=19$) and healthy lung tissue ($n=9$) (Supplementary Table 2 and Fig. 1d). Moreover, positive TKS5 immunostaining was detected in the lungs of IPF/UIP patients ($n=3$), as opposed to control samples ($n=3$), mainly localized in the alveolar epithelium and fibrotic areas (Fig. 1e and Supplementary Fig. 2). Similar conclusions were derived from the analysis of a publicly available single cell RNA sequencing (scRNAseq) dataset of lung tissue from transplant recipients with pulmonary fibrosis ($n=4$) and healthy lung tissue from transplant donors ($n=8$)²². *TKS5* mRNA expression was mostly detected in subsets of epithelial cells, basal cells and especially fibroblasts (Supplementary Fig. 1e, f), where TKS5-expressing LFs were found to belong to a *COL1A1*-expressing subpopulation (Supplementary Fig. 1g, f).

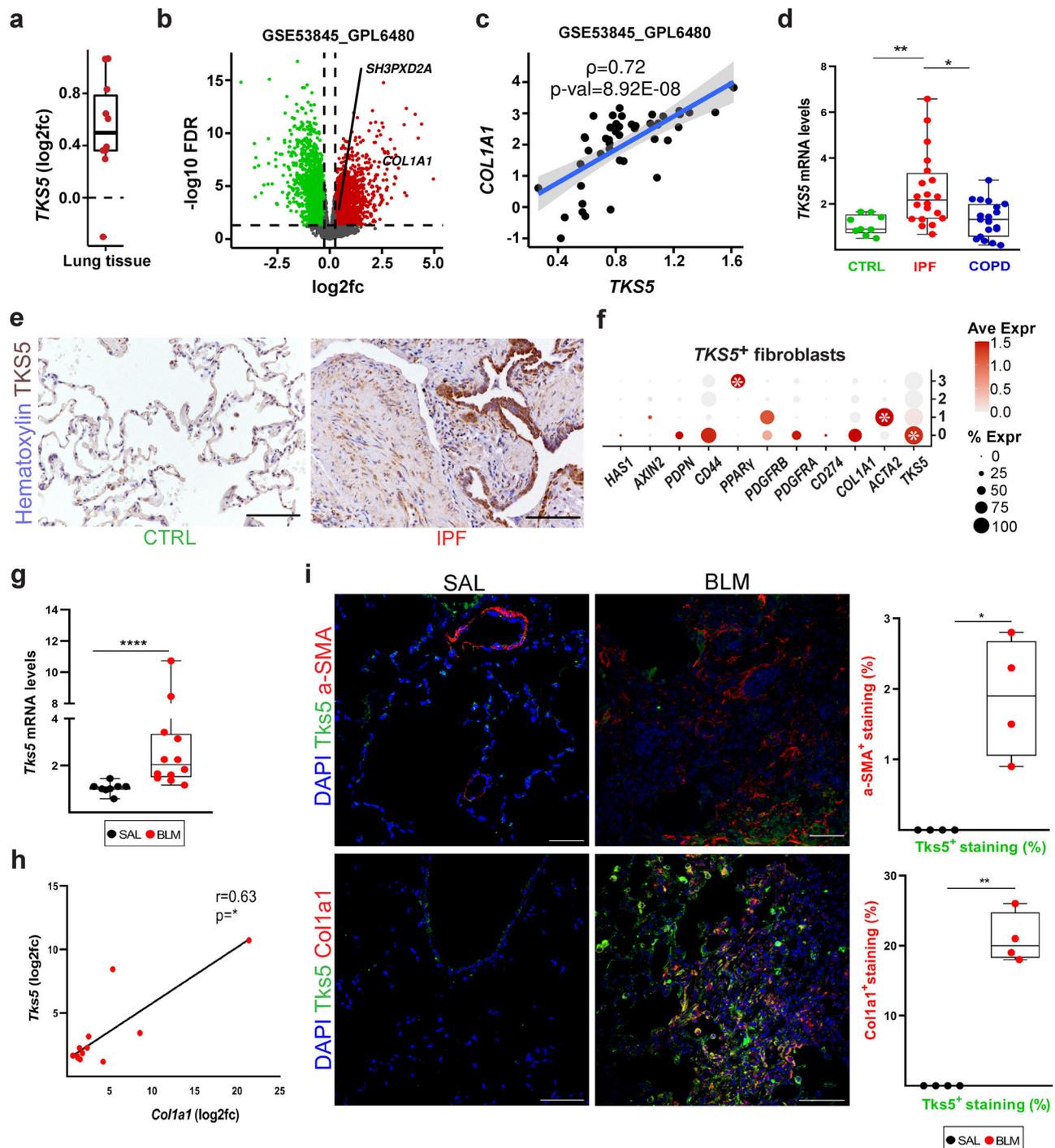
Increased *Tks5* mRNA levels, correlating with *Col1a1* mRNA levels, were also detected in the lung tissue of mice post bleomycin (BLM) administration (Fig. 1g, h), a widely used animal model of pulmonary fibrosis^{23–25} immunostaining localized *Tks5* in the alveolar epithelium and fibrotic areas (Fig. 1i), as in human patients. Moreover, double immunostaining for α SMA or *Col1a1*, prominent activation markers of fibroblasts in both mice and humans, indicated that *Tks5* localized mainly to a *Col1a1* expressing fibroblast subset, as *Tks5* staining overlapped with 20% of *Col1a1* staining, as opposed to a 2% with α SMA staining (Fig. 1i and Supplementary Fig. 3).

Therefore, pulmonary fibrosis in both humans and mice is associated with increased *TKS5* expression, consistently correlated with the expression of *Col1a1*, especially in LFs.

TGF β -induced podosome rosettes is an inherent property of fibrotic LFs

TGF β , among the main pro-fibrotic factors driving disease development in vivo, was found to stimulate *TKS5* mRNA expression in different primary normal human lung fibroblast (NHLF) clones (Fig. 2a), correlating with *COL1A1* mRNA expression (Fig. 2b); identical results were obtained from an independently derived NHLF cell line at a different lab/setting (Supplementary Fig. 4a, b), as well as from the human fibroblastic cell line MRC5 (Supplementary Fig. S4c, d). In agreement with the essential role of TKS5 on podosome formation (colocalization of F-actin with TKS5 or CTTN)²⁶, TGF β , playing a prominent role in proliferation and migration of fibroblasts (Supplementary Fig. 4e–g respectively), potently stimulated the formation of podosomes in NHLFs in vitro, organized in distinctive rosettes (Fig. 2c–f and Supplementary Fig. 5a–f). Moreover, TGF β -induced podosomes in LFs were enriched in MMP9 (Fig. 2g, h and Supplementary Fig. 5g, h), likely contributing to the increased degradation of a fluorescein-conjugated gelatin substrate (Fig. 2i, j), a nominal podosome property.

To examine if the pro-fibrotic milieu in the lungs of IPF patients, which includes TGF β , also stimulate podosome formation in vivo, HLFs from IPF patients (Supplementary Table 3) were cultured in the absence of any stimulation and were stained for podosomes in comparison, under the same conditions (and 7–8 passages), with different NHLF lines derived from healthy tissue. Remarkably, IPF HLFs, irrespective of cell density (Supplementary Fig. 6a), presented with



prominent podosome rosettes (Fig. 3a–d, Supplementary Fig. 6a, b, and Supplementary Movie 1), identical in structure as those stimulated *in vitro* by TGF β , that persist upon prolonged culture *ex vivo*. As shown for TGF β -stimulated NHLFs, IPF HLFs degraded more potently than NHLFs a fluorescein-conjugated gelatin substrate (Fig. 3e, f and Supplementary Fig. 6c).

Phenocopying the human experiments, exposure to TGF β of primary, normal mouse lung fibroblasts (NMLFs) stimulated *Tks5* mRNA expression (Supplementary Fig. 7a), correlating with *Col1a1* expression (Supplementary Fig. 7b), the formation of podosome rosettes (Supplementary Fig. 7c, d) and the degradation of a fluorescein-conjugated gelatin substrate (Supplementary Fig. 7e, f); similar results were obtained with 3T3 embryonic fibroblasts (Supplementary Fig. 7g–k).

Moreover, and as in the case of IPF LFs, mouse primary LFs isolated post-BLM administration presented with increased *Tks5*, *Col1a1* and *Mmp9* expression (Supplementary Fig. 7l–n respectively), exhibiting prominent podosome rosettes in the absence of any stimulation (Supplementary Fig. 7o–p).

Therefore, the pro-fibrotic milieu in the lungs of IPF patients and BLM-treated mice, as well as TGF β , induce TKS5 expression and the formation of podosome rosettes, an inherent fibrotic LF property.

Creation of a series of obligatory and conditional knock out mice for *Tks5*

To enable functional studies on the likely role of *Tks5* in pulmonary fibrosis and pathophysiology in mice, we then created a series of

Fig. 1 | Increased *TKS5* expression in pulmonary fibrosis. **a** *TKS5* mRNA expression in lung tissue from Idiopathic Pulmonary Fibrosis (IPF) patients as compared (log2FC) to controls (CTRL) in different publicly available datasets ($n = 9$) (Supplementary Table 1) at Fibromine. **b** Volcano plot from a representative large dataset ($FC > 1.2$, $FDR < 0.05$). **c** Scatter plot of *TKS5* and *COL1A1* expression in the same dataset with a fitted linear model and 95% CI; correlation was assessed with two-tailed Spearman's test ($\rho > 0.6$; $p = 8.92E-08$). **d** Increased *TKS5* mRNA levels in the lung tissue of IPF (Usual Interstitial Pneumonitis; UIP) patients ($n = 20$) were detected with Q-RT-PCR ($r^2 = 0.98$, $E = 97\%$), as compared with lung tissue from COPD patients ($n = 19$) and control (CTRL) lung tissue isolated from lung cancer patients ($n = 9$) (Supplementary Table 2). Values were normalized to the expression values of the housekeeping gene *B2m* and presented as fold change to CTRL values. Statistical significance was assessed with two-tailed Kruskal-Wallis test ($^{**}p = 0.0076$, $^*p = 0.0129$). **e** Increased *TKS5* immunostaining in fibrotic lungs. Representative images from immunohistochemistry for *TKS5* (brown) in IPF and CTRL lung tissue ($n = 3$; Supplementary Fig. 2); scale bars = 50 μm . **f** *TKS5* is expressed mainly by the *COL1A1*-expressing cluster/LF subpopulation. in a publicly

available scRNAseq dataset (Reyffman, Walter et al. 2019). Statistical significance was assessed with Wilcoxon Rank Sum test ($FC > 1.2$, Bonferroni corrected $p = 8.9E-12 / 1.1E-10 / 2.1E-3$ from left to right). **g, h** *Tks5* and *Col1a1* mRNA expression was interrogated with Q-RT-PCR ($r^2 = 0.89/0.93$; $E = 103\%/96\%$); cumulative result from 3 different experiments. Values were normalized over the expression of *B2m* and presented as fold change (log2) over control ($n = 8/12$). Statistical significance was assessed with two-tailed Mann Whitney test ($^{***}p < 0.0001$). **h** Two tailed spearman correlation plot of *Col1a1* expression in the same samples ($p = 0.0323$; $r = 0.63$). **i** Double immunostaining against *Tks5* (green) and αSMA (Acta2) or *Col1a1* (red); representative images are shown, followed by their respective quantification ($n = 4$) with Image J; scale bars=50 μm ; a representative experiment out of 3 successful independent ones are shown. Statistical significance was assessed with two-tailed Welch's test ($p = 0.0211$, $^{**}p = 0.0013$). In all panels all samples are biologically independent; boxplots visualize the median of each distribution; upper/lower hinges represent 1st/3rd quartiles; whiskers extend no further than 1.5 * IQR from the respective hinge. Source data for all panels are provided as a Source Data file.

obligatory and conditional knock out mice for *Tks5* (*Sh3pxd2a*). The *Sh3pxd2a* locus has been already targeted by the European Conditional Mouse Mutagenesis Program (EUCOMM), aiming to knock out all mouse genes in a high throughput approach²⁷. In this context, the exon 11 of the *Sh3pxd2a* gene was loxP-flanked, while a LacZ/neomycin reporter/selection cassette was placed upstream, including two FRT sites; this allele is referred to as “targeted mutation 1a” (tm1a; Supplementary Fig. 8a)²⁷. The targeted ES cells were then microinjected into C57Bl/6N blastocysts by the Wellcome Trust Sanger Institute (WTSI), that were transferred in pseudopregnant females to yield the *Sh3pxd2a*^{tm1a(EUCOMM)Wtsi/+} heterozygous mice (Supplementary Fig. 8a). Frozen sperm of these mice was obtained from WTSI, via the INFRAFROntier [<https://www.infrafrontier.eu/>] consortium^{28,29} and the European Mutant Mouse Archive- EMMA [<https://www.infrafrontier.eu/emma/emma-services/?keyword=sh3pxd2a&category=strains>]), that was directly injected to mice in the transgenic facility of “BSRC Fleming” via IVF technology to yield the *Sh3pxd2a*^{tm1a(EUCOMM)WtsiFimg/+} heterozygous mice. Mice were genotyped following the corresponding strategy from EUCOMM, that queries three different genomic fragments (lacZ, WT allele, tm1a allele) by performing three independent PCR reactions (Supplementary Fig. 8b, c). Moreover, the successful targeting was also verified with long range PCR for both the 5' and 3' arms flanking the floxed region with primers against inserted sequences (Supplementary Fig. 8d, e).

To obtain the tm1b reporter allele (Supplementary Fig. 8a) *Sh3pxd2a*^{tm1a/Fleming/+} mice were mated with transgenic mice expressing the Cre recombinase under the control of the Cytomegalovirus (CMV) promoter in all mouse tissues and cells (*Tg-CMV-Cre*)³⁰. Genetic recombination of the obtained *Sh3pxd2a*^{tm1b(EUCOMM)WtsiFimg/+} mice was verified with genomic PCR (Fig. S7B, C). Q-RT-PCR in lung tissues indicated a 50% reduction of *Sh3pxd2a* mRNA levels indicating proper gene targeting (Supplementary Fig. 8g). X-gal staining, detecting LacZ expression from the promoter of *Tks5* (Supplementary Fig. 8a, Tm1b), localized transcriptional *Tks5* activation (throughout development, neonatal and adult life) mainly in arterial endothelium of the lung (Supplementary Fig. 8i). No obvious gross macroscopic abnormalities were observed, while heterozygous mice were healthy and fertile.

The haploinsufficient *Sh3pxd2a*^{tm1b(EUCOMM)WtsiFimg/+} and *Sh3pxd2a*^{tm1d(EUCOMM)WtsiFimg/+} (*Tks5*^{+/-}) mice, presented with a 50% reduction of *Tks5* mRNA levels in the lung (Supplementary Fig. 8g-h), while X-gal staining (in the reporter tm1b strain) localized transcriptional *Tks5* activation (throughout development, neonatal and adult life) mainly in endothelial and smooth muscle cells in healthy conditions (Supplementary Fig. 8i). No obvious gross macroscopic abnormalities were observed, while heterozygous mice were healthy and fertile.

Intercrossing of heterozygous mice *Tks5*^{+/-} mice yielded no homozygous knockout mice, indicating that *Tks5* has an essential role

in mouse development, as previously reported for an obligatory knock out strain¹⁹.

A similar genetic strain, *Sh3pxd2a*^{tm1b(EUCOMM)Wtsi/+} was created by WTSI from the *Sh3pxd2a*^{tm1a(EUCOMM)Wtsi/+} mice via a cell permeable HTN-Cre³¹. *Sh3pxd2a*^{tm1b(EUCOMM)Wtsi/+} heterozygous mice were systematically phenotyped from the INFRAFROntier consortium^{28,29} on our behalf, following a relative competitive call. Results indicated that *Sh3pxd2a*^{tm1b(EUCOMM)Wtsi/+} mice present with no major pathophysiological abnormalities, apart from an increase of serum alkaline phosphatase in females (measurements.chart[<https://www.mousephenotype.org/data/genes/MGI:1298393#phenotypes-section>]). Moreover, a viability primary screen phenotypic assay was performed on the novel mutant strain by WTSI (data[chart[[https://www.mousephenotype.org/data/charts?accession=MGI:1298393&allele_accession_id=MGI:5636944&zygosity=homozygote¶meter_stable_id=IMPC_VIA_001_001&pipeline_stable_id=MGP_001&procedure_stable_id=IMPC_VIA_001_001¶meter_stable_id=IMPC_VIA_001_001&phenotyping_center=WTSI](https://www.mousephenotype.org/data/charts?accession=MGI:1298393&allele_accession_id=MGI:5636944&zygosity=homozygote¶meter_stable_id=IMPC_VIA_001_001&pipeline_stable_id=MGP_001&procedure_stable_id=IMPC_VIA_001¶meter_stable_id=IMPC_VIA_001_001&phenotyping_center=WTSI)]]), confirming the requirement of *Tks5* for embryonic development.

Moreover, and to generate the conditional tm1c allele (Fig. S7A), *Sh3pxd2a*^{tm1a(EUCOMM)WtsiFimg/+} mice were crossed with transgenic mice expressing the Flp recombinase under the control of the Cytomegalovirus (CMV) promoter in all mouse tissues and cells (*Tg-CMV-Flp*)³². Genetic recombination of the obtained *Sh3pxd2a*^{tm1c(EUCOMM)WtsiFimg/+} mice was verified with genomic PCR (Supplementary Fig. 8c). To generate the conditional tm1d allele (Supplementary Fig. 8a), *Sh3pxd2a*^{tm1c(EUCOMM)WtsiFimg/+} mice were crossed with transgenic mice expressing the Cre recombinase under the control of the Cytomegalovirus (CMV) promoter in all mouse tissues and cells (*Tg-CMV-Cre*)³⁰. Genetic recombination of the obtained *Sh3pxd2a*^{tm1d(EUCOMM)WtsiFimg/+} mice was verified with genomic PCR (Supplementary Fig. 8c). Q-RT-PCR in lung tissue indicated a 50% reduction of *Sh3pxd2a* mRNA levels indicating proper gene targeting (Supplementary Fig. 8h).

***Tks5* haploinsufficiency in mice attenuates BLM-induced pulmonary fibrosis**

To genetically dissect the likely role of *Tks5* in pulmonary fibrosis, BLM was administered to 8-10-week-old C57Bl6/J *Tks5*^{+/-} mice and WT littermates (Fig. 4a, b), which were sacrificed 14 days post BLM (at the peak of the disease in the local settings), as previously described²⁴. No weight loss, an overall systemic health indicator, was observed in *Tks5*^{+/-} mice (Fig. 4c), as opposed to wt mice. Vascular leak and pulmonary edema were significantly reduced in *Tks5*^{+/-} mice, as indicated by the total protein concentration in the bronchoalveolar lavage fluid (BALF), determined with the Bradford assay (Fig. 4d). Inflammatory cells in the BALF, as measured by hemacytometer, were found

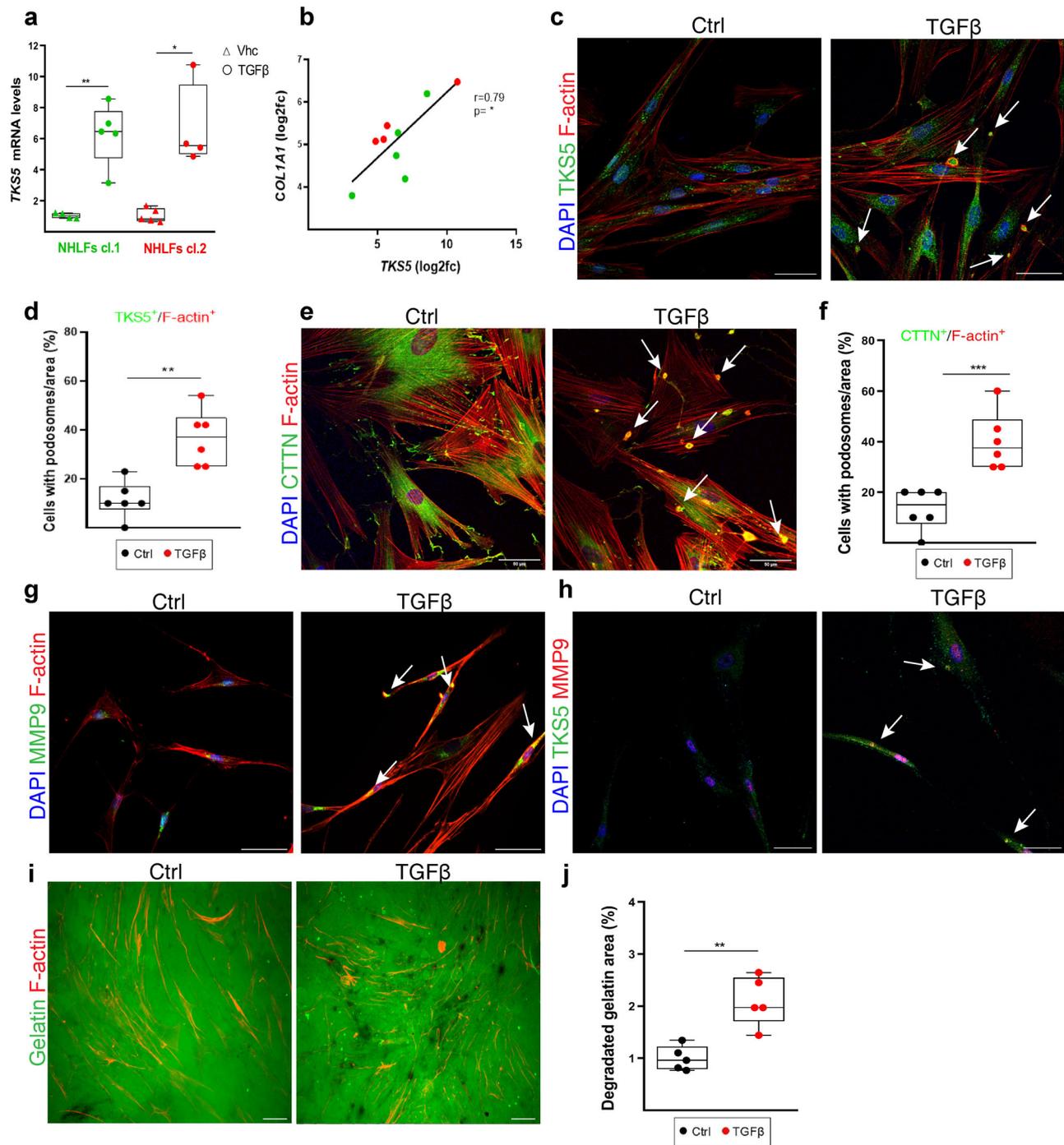


Fig. 2 | TGFβ induces the formation of podosome rosettes in normal human lung fibroblasts (NHLFs). Serum starved, sub-confluent (70–80%), primary NHLFs were stimulated with recombinant human TGFβ (10 ng/ml) for 24 h; a representative experiment out of 4 successful independent ones is shown. **a**, **b** TKS5 and COL1A1 mRNA expression was interrogated with Q-RT-PCR ($r^2 = 0.94/0.92$; $F = 98.3\%/93\%$ respectively) in two NHLF clones (cl.1, cl.2). Values were normalized to the expression values of the housekeeping gene *B2M* and presented as fold change over control; $n = 4/5/4/4$; statistical significance was assessed with two-tailed Welch's test (a/cl.1) and two-tailed Mann Whitney test (a/cl.2); $^*p = 0.0012$, $^*p = 0.0159$ respectively. **b** Two tailed Pearson correlation plot of COL1A1 expression in the same samples ($p = 0.0116$; $r = 0.79$). **c–j** Representative composite images from double immunostaining, and respective quantifications, for: **c** F-actin/

TKS5 (red/green), **e** F-actin/Cortactin (CTTN; red/green), **g** F-actin/MMP9 (red/green), **h** TKS5/MMP9 (green/red). Cells are counterstained with DAPI; scale bars 50 μm; arrows indicate representative podosomes; separate images and proof of colocalization of signals can be found at Supplementary Fig. 5. **d**, **f**. Quantification of the number of podosome-containing cells per optical field ($n = 6$); statistical significance was assessed with two-tailed t-test; $^*p = 0.0011$, $^{***}p = 0.0009$.

i Representative images of the TGFβ-induced degradation (black holes) of a fluorescein-conjugated gelatin substrate. **j** Quantification of gelatin degradation area, normalized to control ($n = 5$); statistical significance was assessed with two-tailed t-test; $^*p = 0.0016$. Source data for all relative panels (a, b, d, f, j) are provided as a Source Data file.

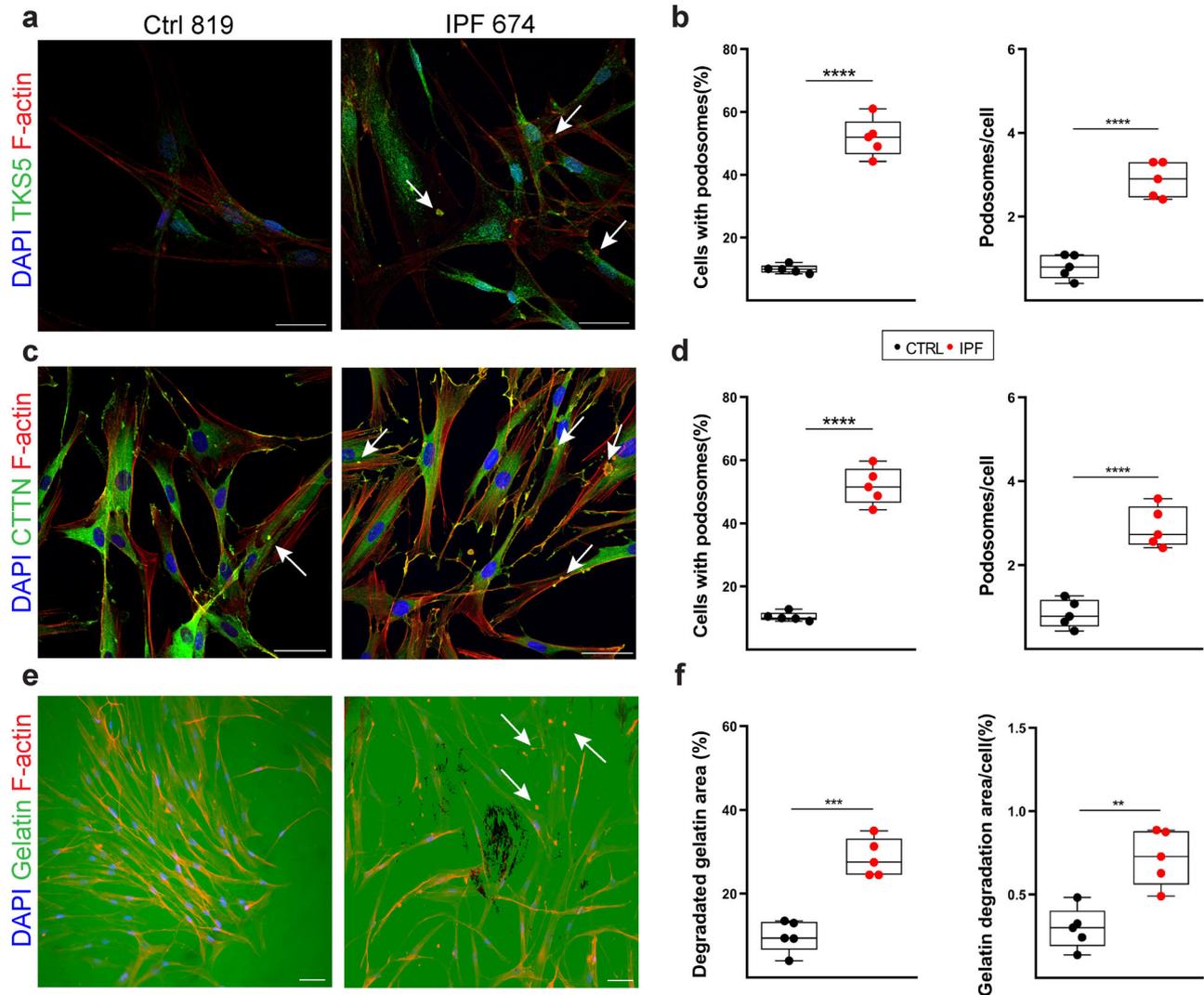


Fig. 3 | The formation of extracellular matrix (ECM) degrading podosome rosettes is an inherent property of IPF human lung fibroblasts (HLFs). Serum starved, sub-confluent (70–80%), primary IPF-HLFs and normal HLFs (NHLFs) were immunostained for F-actin (red) and (a) TKS5 (green) or (c) cortactin (CTTN; green) and counter stained with DAPI (blue); $n = 5$; scale bars = 50 μm . Representative images from representative clones are shown. **b, d** Cumulative quantification of the number of podosome-containing cells (%) and the number of podosomes per cell per optical field respectively. Statistical significance was assessed with two-tailed t -test (**b**) or two-tailed Welch's test (**d**) ($^{***}p < 0.0001$). **e** The same clones were cultured on a fluorescein-conjugated gelatin substrate (green) and were stained for

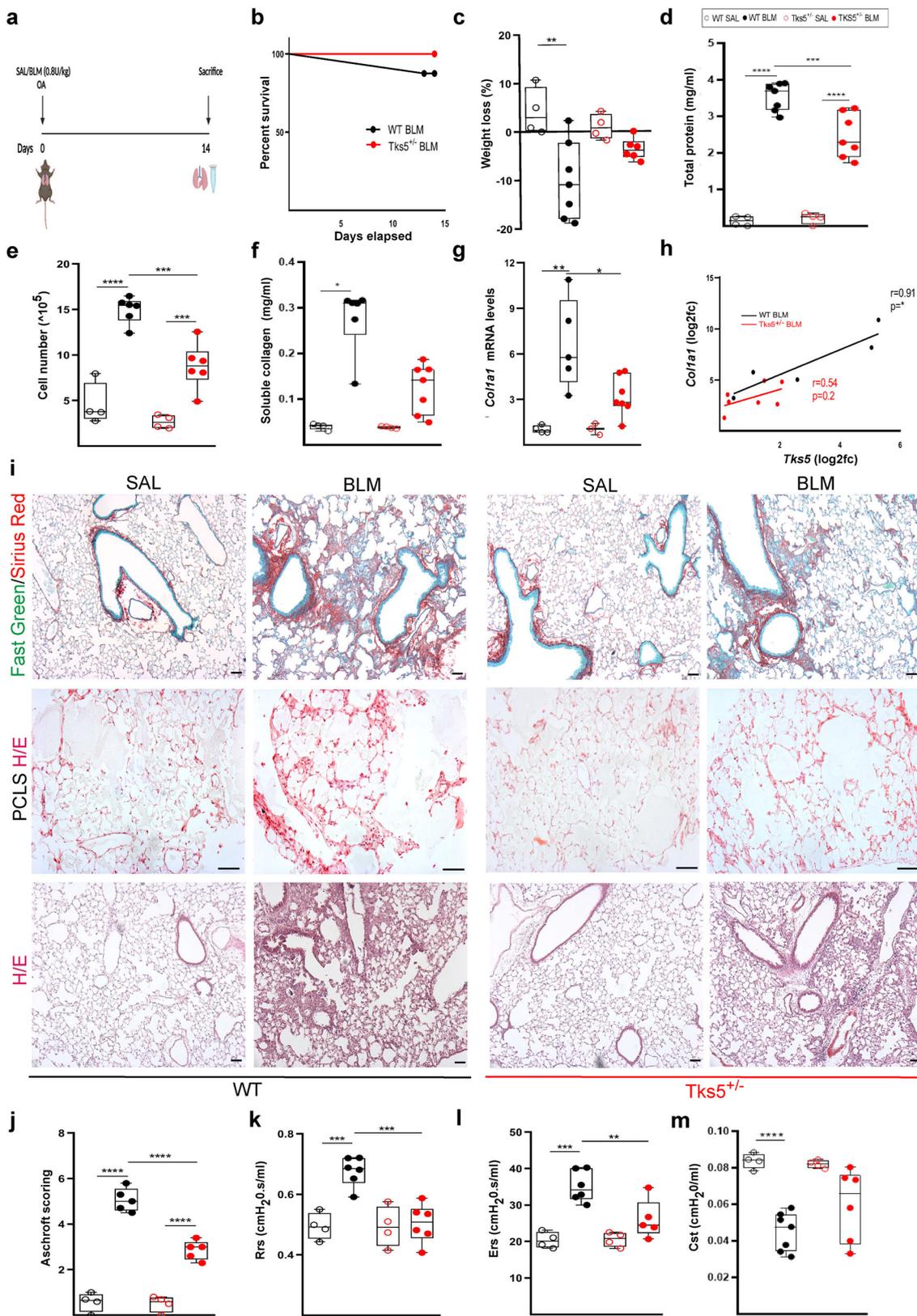
F-actin (red) and counter stained with DAPI (blue); representative images are shown. **f** Quantification of the percentage of the degraded gelatin for all clones cumulatively, and the quantification of gelatin degradation area per cell, as quantified with ImageJ; statistical significance was assessed with two-tailed t -test; $^{***}p = 0.0001$, $^{**}p = 0.0020$; additional clones and controls are shown at Supplementary Fig. 6. In all panels all samples are biologically independent; boxplots visualize the median of each distribution; upper/lower hinges represent 1st/3rd quartiles; whiskers extend no further than 1.5 IQR from the respective hinge. Source data for all relative panels (**b, d, f**) are provided as a Source Data file.

significantly reduced in $Tks5^{+/-}$ mice (Fig. 4e); so were soluble collagen BALF levels, as determined by the Sirius red assay (Fig. 4f), in concordance with *Coll1a1* mRNA expression in the lung tissue from the same mice, as determined with Q-RT-PCR (Fig. 4g, h). Histological analysis revealed decreased collagen deposition in $Tks5^{+/-}$ mice post BLM, as quantified by Sirius red/Fast green staining (Fig. 4i), and fewer peribronchiolar and parenchymal fibrotic regions were detected (Fig. 4i), as reflected in the Ashcroft score (Fig. 4j); similar conclusions were drawn upon the histological evaluation of Precision Cut Lung Slices (PCLS) prepared from the same mice and cultured ex vivo (Fig. 4i). The relative protection from the BLM-induced tissue architecture distortion upon the genetic reduction of *Tks5* expression was also reflected in lung respiratory functions, as measured with FlexiVent (Fig. 4k–m). Therefore, *Tks5* expression, and likely the formation of podosomes, were shown to

have a major role in BLM-induced pulmonary fibrosis, and therefore likely IPF.

***Tks5* haploinsufficiency in mouse LFs decreases ECM-regulated podosome formation and ECM invasion**

To functionally dissect the relative protection of $Tks5^{+/-}$ mice from BLM-induced pulmonary fibrosis, primary LFs were isolated from littermate wt and $Tks5^{+/-}$ mice and were exposed to TGF β , as before. $Tks5^{+/-}$ LFs, expressing ~50% of *Tks5* (Fig. 5a), presented with decreased numbers of podosomes in response to TGF β (Fig. 5b, c), reaffirming the seminal role of *Tks5* in podosome formation²⁶, as well as with decreased proliferation (24 h; Fig. 5d). As podosomes are known to promote ECM invasion, we then examined the ability of LFs to invade acellular ECM (aECM) prepared from the lungs of mice (Supplementary Fig. 9a, b), in a transwell invasion chamber (6 h; Fig. 5e). The reduction of podosomes



was associated with a decreased TGF β -induced invasion of *Tks5*^{+/-} LFs in aECM (Fig. 5f). Moreover, reaffirming in mice the inherent character of podosome formation in LFs, post BLM *Tks5*^{+/-} LFs presented with reduced numbers of podosomes in comparison with wt LFs isolated from littermate mice (Fig. 5g, h), resulting in defective aECM invasion (Fig. 5i). Therefore, the in vivo demonstrated pathogenic role of *Tks5* in

pulmonary fibrosis includes the formation of podosomes in LFs and the promotion of their ECM invasion.

To obtain additional mechanistic insights, wt and *Tks5*^{+/-} LFs were exposed to TGF β , as above, and their global expression profile was interrogated with 3' UTR RNA sequencing (Quant-Seq LEXO-GEN). Differential expression analysis between TGF β -induced *Tks5*^{+/-}

Fig. 4 | *Tks5* haploinsufficiency in mice attenuates bleomycin (BLM)-induced pulmonary fibrosis (PF). **a** Schematic presentation of the BLM-induced PF model. **b** Kaplan Meyer survival curve post BLM administration. **c** Weight change post BLM administration; $n = 4/7/4/6$. Statistical significance was assessed with two-tailed one-way ANOVA; $^{***}p = 0.031$. **d** Total protein concentration in bronchoalveolar lavage fluids (BALFs), as determined with the Bradford assay; $n = 4/7/4/7$. Statistical significance was assessed with two-tailed one-way ANOVA; $^{****}p < 0.0001$, $^{***}p = 0.0009$. **e** Inflammatory cell numbers in BALFs, as counted with a hemacytometer; $n = 4/6/4/6$. Statistical significance was assessed with two-tailed one-way ANOVA; $^{***}p < 0.0001$, $^{**}p = 0.0002/0.0008$. **f** Soluble collagen levels in the BALFs were detected with the direct red assay; $n = 4/6/4/7$. Statistical significance was assessed with two-tailed Kruskal Wallis; $^{*}p = 0.0124$. **g, h** *Tks5* and *Col1a1* mRNA expression was interrogated with Q-RT-PCR; $n = 4/5/3/7$; values were normalized over the expression of the housekeeping gene *B2m* and presented as fold change over control. Statistical significance was assessed with two-tailed one-way ANOVA; $^{**}p = 0.0012$, $^{*}p = 0.0207$. **h** Two tailed pearson correlation plot of *Col1a1* expression in the same samples; $^{*}p = 0.0342$; $r = 0.91/0.54$. **i** Representative images from lung

sections of murine lungs of the indicated genotypes, stained with Fast Green/Sirius Red (green/red; first row), from Hematoxylin & Eosin (H&E)-stained Precision cut lung slices (PCLS) (second row) and H&E-stained lung sections (third row); scale bars 50 μm . **j** Quantification of fibrosis severity in H/E stained lung sections via Ashcroft scoring; $n = 4/5/4/5$. Statistical significance was assessed with two-tailed one-way ANOVA; $^{****}p < 0.0001$. **k** Rrs, mean respiratory system resistance as measured with Flexivent; $n = 4/6/4/6$. Statistical significance was assessed with two-tailed one-way ANOVA; $^{***}p = 0.0008/0.0004$. **l** Ers, mean respiratory system elastance as measured with Flexivent; $n = 4/6/4/5$. Statistical significance was assessed with two-tailed one-way ANOVA; $^{**}p = 0.0002$, $^{*}p = 0.0081$. **m** Cst, mean static lung compliance as measured with Flexivent; $n = 4/7/4/6$. Statistical significance was assessed with two-tailed one-way ANOVA followed by Welch's correction; $^{***}p < 0.0001$. In all panels cumulative results from 2 different experiments are shown; all samples are biologically independent; boxplots visualize the median of each distribution; upper/lower hinges represent 1st/3rd quartiles; whiskers extend no further than 1.5 IQR from the respective hinge. Source data for all relative panels (**a-h, j-m**) are provided as a Source Data file.

and wt LFs, revealed 3648 differentially expressed genes (DEGs; FC > 1.2, FDR corr. $p < 0.05$; Supplementary Data 1 and Supplementary Fig. 10a); among them 418 DEGs have been previously associated with pulmonary fibrosis, as detected with text mining of abstract co-occurrence of identified DEGs with fibrosis keywords (Supplementary Data 1). *Stat1*, *Cebpa*, and *Ar* transcription factors (TFs), where found downregulated in *Tks5*^{-/-} LFs along with several of their target genes (Supplementary Data 1 and Supplementary Fig 10b). Gene set enrichment analysis (GSEA) performed on DEGs revealed that the most affected cellular components (CC), molecular functions (MF) and biological processes (BP) all relate to the ECM (Fig. 6a and Supplementary Data 2). "Collagen containing ECM" (GO:0062023) was most prominent due to the down regulation of several ECM related genes such as collagens and MMPs/TIMPS/Adamts (Fig. 6b and Supplementary Fig. 10c). In this context and given the observed consistent correlation of *Tks5* and *Col1a1* expression, *Tks5*^{-/-} LFs post BLM, containing fewer podosomes and exhibiting defective aECM invasion (Fig. 5g-i), were found to produce significantly less *Col1a1* (Fig. 6c). Vice versa, culture of primary NMLFs on *Col1a1*-rich aECM prepared from the lungs of mice post BLM (Supplementary Fig. 9c), stimulated *Tks5* expression (Fig. 6d, e) and the formation of podosomes (Fig. 6f), and further stimulated *Col1a1* expression (Fig. 6g), indicating an ECM-podosome cross talk in the perpetuation of LF activation.

Src-inhibition potently reduces podosome formation and attenuate pulmonary fibrosis

To identify pharmaceutical compounds that can induce a similar transcriptional profile as that of the defective in ECM invasion *Tks5*^{-/-} LFs, the TGF β -induced *Tks5*^{-/-} LFs profile was queried against the connectivity map (CMap) LINCSL1000 database (Fig. 7a), a public resource that contains >10⁶ gene expression signatures of different cell types treated with a large variety of small molecule compounds³³. Among the identified compounds with similar expression signatures, several have already been shown to have a positive effect in disease pathogenesis in animal models (Fig. 7a and Supplementary Table 4). The identified possible therapeutic targets include the PDGF and VEGF receptors, which are pharmacologically targeted by the current IPF standard of care (SOC) compound nintedanib³⁴. More importantly, the list also includes an inhibitor of Src, a TGF β /PDGF-inducible, non-receptor tyrosine kinase essential for podosome formation³⁵. To verify the in silico findings in our experimental settings, TGF β -activated NHLFs were incubated with nontoxic, increasing concentrations of nintedanib and A-419259, a commercially available src inhibitor. Both nintedanib but especially A-419259 reduced both *TKSS* and *COL1A1* expression (Fig. 7b-e), as well as podosome formation (Fig. 7f, g) and aECM LF invasion (Fig. 7h).

To examine possible therapeutic effects of src inhibition in pulmonary fibrosis, we generated mouse precision cut lung slices (PCLS) post BLM (d11) administration, which were then incubated with A-419259 for 3 consecutive days, resulting in the attenuation of pulmonary fibrosis (Fig. 7i). Moreover, the same inhibitor was administered for 6 days by inhalation (4 ml of 0182 mg/ml for 5 mins/6 mice, corresponding to 2 mg/Kg per mouse) to conscious, softly re-strained mice, in a therapeutic mode (7d post BLM; Fig. 8a); no lethality was observed (Fig. 8b); minimal changes were observed in weight loss (Fig. 8c). Remarkably, src inhibition decreased, pulmonary edema (Fig. 8d) and inflammation (Fig. 8e), and attenuated Col production (Fig. 8f, g). Accordingly, src inhibition attenuated collagen deposition in the lung tissue, and prevented BLM-induced architectural distortion (Fig. 8h, i). Therefore, the *TKSS*-mediated podosome formation is a druggable pathologic process, which can be potently targeted by Src inhibition.

Discussion

Increased *TKSS* expression was detected, for the first time in a non-malignant disease¹⁵, in the lung tissue of IPF patients and BLM-treated mice (Fig. 1 and Supplementary Figs. 1-2). Increased *TKSS* expression has been previously reported, beyond cancer cell lines, in lung adenocarcinoma²⁰, further extending the similarities of IPF and lung cancer¹³. *TKSS* mRNA expression in the lung tissue, of both humans and mice, correlated with the mRNA expression of *COL1A1*, a hallmark of deregulated expression in IPF, while *TKSS* expression in fibrotic lungs was predominantly localized in the alveolar epithelium and *COL1A1*-expressing LFs (Fig. 1 and Supplementary Figs. 1-3), pending larger scale immunohistochemical studies.

TGF β , the prototypic pro-fibrotic factor, was found to be a very potent inducer of *TKSS* expression and podosome formation in fibroblasts (NHLFs, NMLFs, MRC5, 3T3)(Fig. 2, Supplementary Figs. 4-5), as previously reported only for THP-1 macrophages^{36,37} and primary aortic endothelial cells^{38,39}. Other well established pro-fibrotic growth factors in the lung have been reported to modulate podosome formation in different cell types: PDGF in synovial fibroblasts⁴⁰ and smooth muscle cells⁴¹, and VEGF in endothelial cells⁴², suggesting that they could exert similar stimulatory effects on LFs. Moreover, PGE₂, which suppresses pulmonary fibrosis⁴³, has been reported to promote the dissolution of podosomes in dendritic cells⁴⁴, suggesting that the diminished PGE₂ levels in IPF⁴³ also favor the formation of podosomes. Remarkably, the formation of podosomes in LFs was shown to be an inherent property of IPF and post BLM LFs that can be maintained in culture in the absence of any stimulation (Fig. 3, Supplementary Figs. 6-7). In agreement, increased invadosome formation was very recently reported in IPF LFs (3-7 passages), correlating with fibrosis severity⁴⁵. Therefore, podosome formation is an unappreciated central

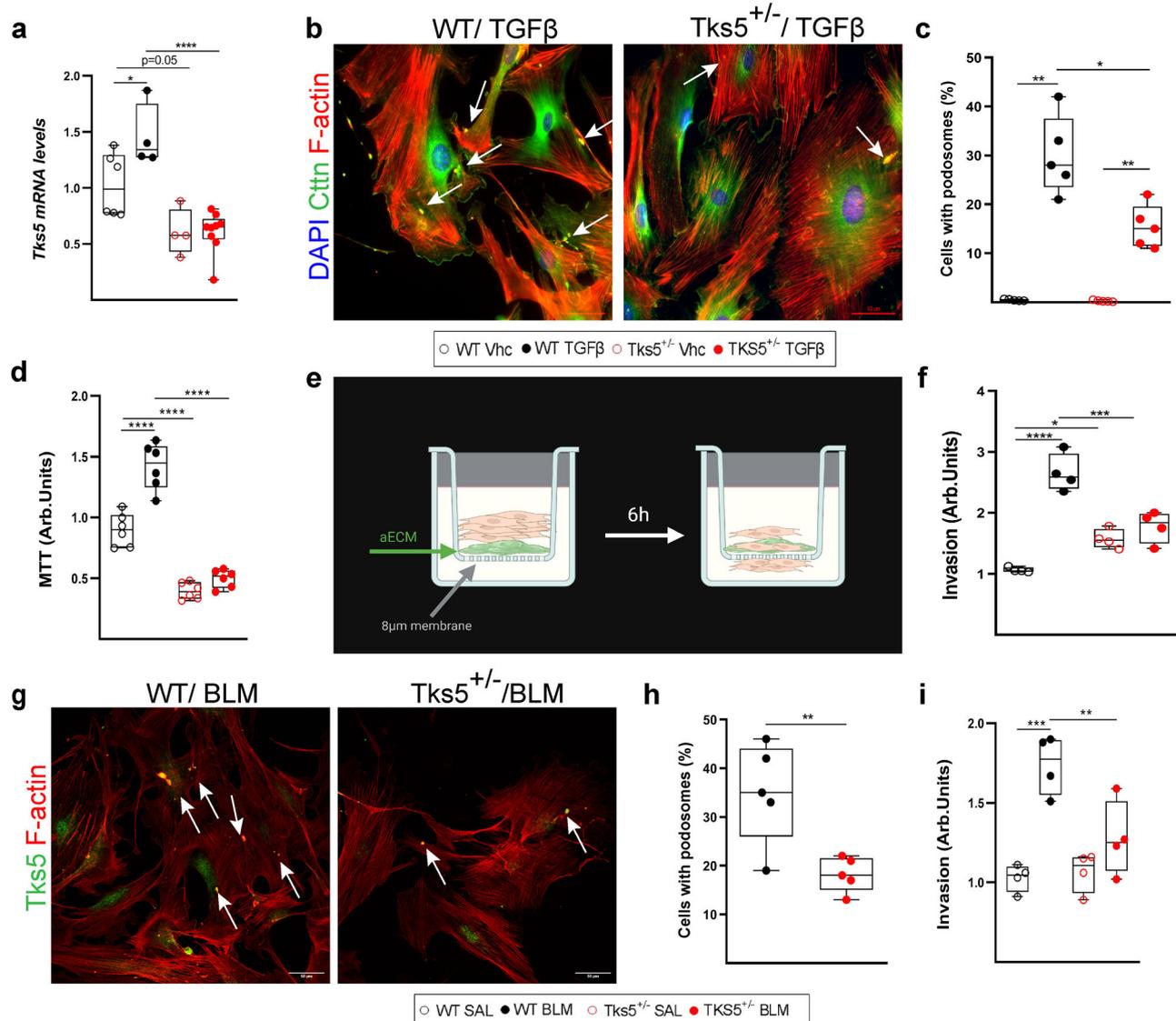


Fig. 5 | Tks5 haploinsufficiency in mouse lung fibroblasts (LFs) decreases podosome formation and extracellular matrix (ECM) invasion. Serum starved primary normal mouse LFs (NMLFs) from WT and Tks5^{+/-} mice were stimulated with recombinant TGF-β1 (10 ng/ml for 24 h). **a** *Tks5* mRNA expression was interrogated with Q-RT-PCR; $n = 6/4/4/9$. Values were normalized over the expression of the housekeeping gene *B2m* and presented as fold change over control. Statistical significance was assessed with two-tailed one-way ANOVA; * $p = 0.0464$, **** $p < 0.0001$. **b** Representative composite images from double immunostaining for F-actin (red) and Cortactin (Ctnn; green) counter stained with DAPI (blue); arrows indicate representative podosomes. **c** Quantification of the number of podosome-containing cells per optical field; $n = 5$. Statistical significance was assessed with two-tailed one-way ANOVA followed by Welch's correction; * $p = 0.004/0.0053$ $p = 0.0411$. **d** TGFβ-induced NMLFs proliferation was assessed with the MTT assay; $n = 6$. Statistical significance was assessed with two-tailed one-way ANOVA; **** $p < 0.0001$. **e** Schematic presentation (biorender.com) of LFs invasion into aECM, upon TGFβ stimulation. After 6 h, cells that had invaded into the lower surface of

the upper chamber were stained, lysed and absorbance values were measured. **f** Invasion capacity of NMLFs, upon TGF-β stimulation ($n = 4$), as detected with the transwell invasion assay. Statistical significance was assessed with two-tailed one-way ANOVA; **** $p < 0.0001$, * $p = 0.0266$, *** $p = 0.0005$. **g** Representative composite images from double immunostaining for F-actin (red) and Tks5 (green) in NMLFs isolated from WT and Tks5^{+/-} mice, post bleomycin (BLM) administration; arrows indicate representative podosomes; scale bars 50 μm. **h** Quantification of the number of podosome-containing cells per optical field ($n = 5$). Statistical significance was assessed with two-tailed t-test; * $p = 0.009$. **i** Invasion capacity of LFs post BLM, as detected with the transwell invasion assay; $n = 4$. Statistical significance was assessed with two-tailed one-way ANOVA; **** $p = 0.0003$ ** $p = 0.0099$. In all panels, all samples are biologically independent; boxplots visualize the median of each distribution; upper/lower hinges represent 1st/3rd quartiles; whiskers extend no further than 1.5 IQR from the respective hinge. Source data for all panels are provided as a Source Data file.

response of LFs to pro-fibrotic factors and a major inherent characteristic of IPF LFs, likely contributing to their accumulation and the formation of foci, a hallmark of UIP/IPF.

Col1a1 mRNA levels were found to consistently correlate with *Tks5* mRNA levels in both humans and mouse lung tissue or LFs (Figs. 1–3 and 5, and Supplementary Figs. 1, 3, and 9). Tks5^{+/-} LFs were shown to produce less Col1a1 post BLM (Fig. 6c), as also reflected in the reduced overall collagen deposition in the lungs of Tks5^{+/-} mice (Fig. 4g). Culture

of LFs on Col1a1-rich aECM promoted *Tks5* expression and podosome formation (Fig. 6d–), as well as further *Col1a1* expression, emphasizing a TGFβ-induced Col1a1-podosomes interdependency in the context of the suggested crosstalk of ECM with podosomes⁴⁶. Accordingly, Col1 has been shown to stimulate Tks5-dependent growth, while the degree of collagen fibrilization has also been reported to have a decisive effect on podosome formation⁴⁷, likely through the Discoidin domain receptors (DDR) that mediate collagen binding⁴⁸. The expression of both

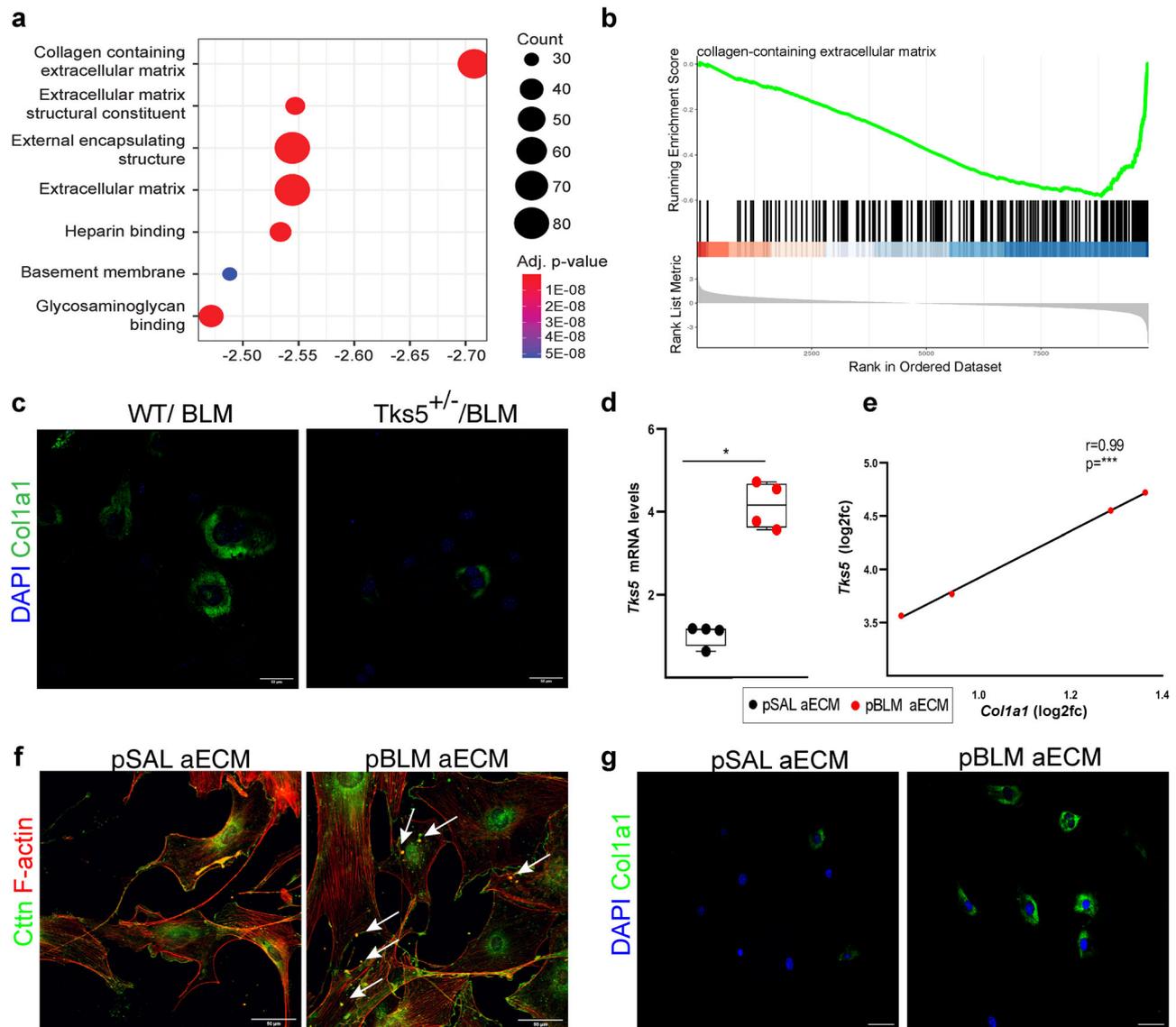


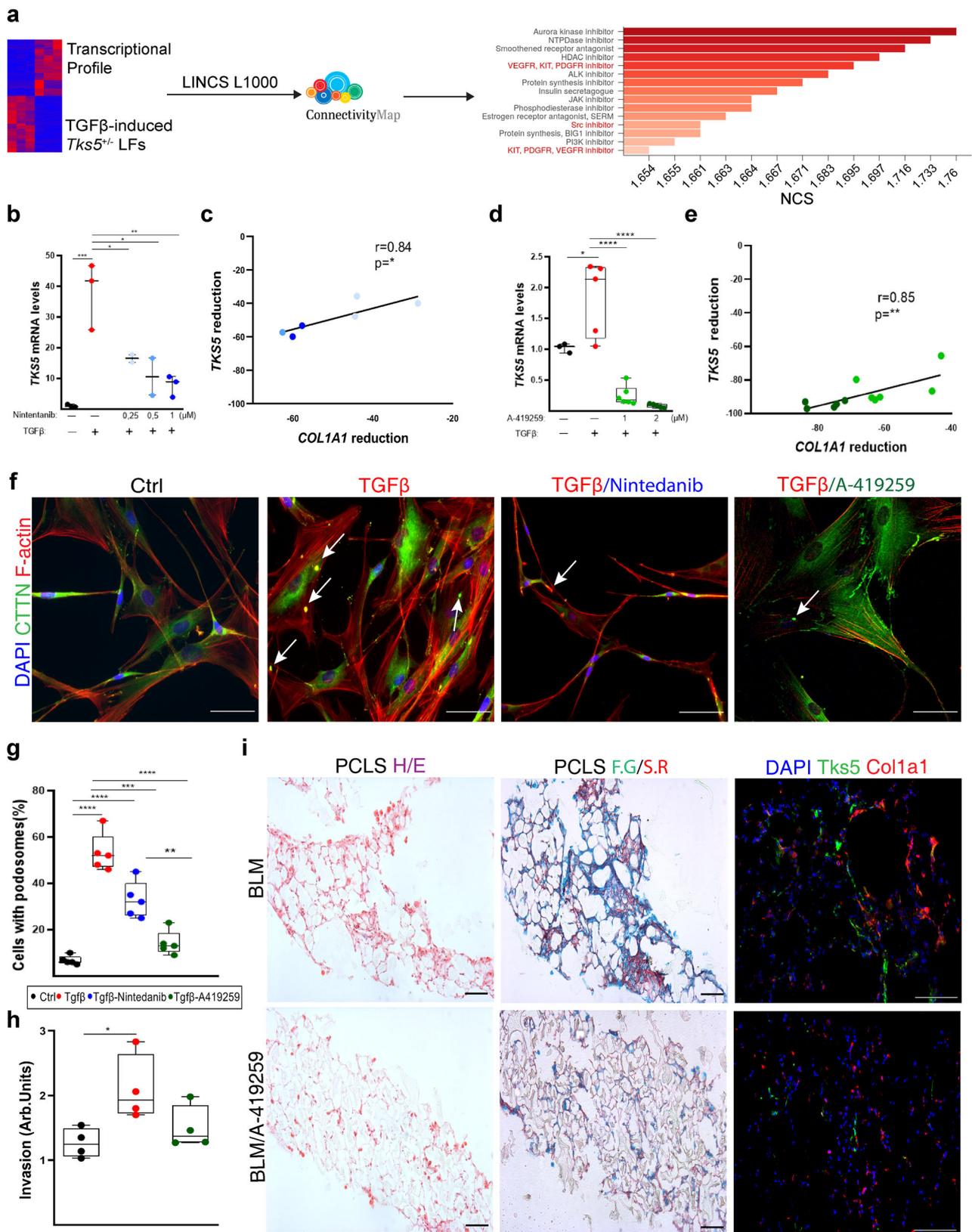
Fig. 6 | *Tks5* haploinsufficiency in mouse lung fibroblasts (LFs) disrupts extracellular matrix (ECM) homeostasis, that critically controls podosome formation and ECM invasion. **a** ECM-related gene ontology components are enriched for genes down-regulated in TGF β stimulated $Tks5^{+/-}$ LFs compared to their WT TGF β treated counterparts. Gene-set enrichment analysis (GSEA) on expression data pre-ranked according to their fold change values. **b** Collagen containing extracellular matrix is the term most enriched in down-regulated genes according to gene-set enrichment analysis (GSEA). **c** Serum starved WT and $Tks5^{+/-}$ LFs were immunostained for Col1a1 (green) and counter stained with DAPI (blue). Representative images are shown; scale bars = 50 μ m. **d**, **e** Serum starved WT primary LFs were cultured in post saline (pSAL) and post bleomycin (pBLM) aECM. *Tks5* and *Col1a1* mRNA expression was interrogated with Q-RT-PCR; $n = 4$; values

were normalized over the expression of the housekeeping gene *B2m* and presented as fold change over control. Statistical significance was assessed with two-tailed Mann-Whitney test; $p = 0.0286$. **e** Two tailed pearson correlation plot of *Col1a1* expression in the same samples; $**p = 0.0004$, $r = 0.99$. **f**, **g** Representative composite images from double immunostaining for F-actin (red) and Cortacin (Ctnn; green; **f**) or Col1a1 (**g**; green) counter stained with DAPI (blue); arrows indicate representative podosomes; scale bars = 50 μ m. In all panels, representative experiment out of 2 successful independent ones are shown; all samples are biologically independent; boxplots visualize the median of each distribution; upper/lower hinges represent 1st/3rd quartiles; whiskers extend no further than 1.5 IQR from the respective hinge. Source data for all panels are provided as a Source Data file.

DDR1 and 2 were found downregulated in TGF β -induced $Tks5^{+/-}$ LFs (Supplementary Data 1), suggesting that DDR signaling, and consequent tyrosine kinase activation could mediate the observed TGF β -Col1a-induced podosome formation. While the reverse signaling from the overproduction of collagen and the regulation of DDR1/2 should be further researched in the context of IPF, the therapeutic potential of DDR inhibitors, currently explored for metastatic cancer⁴⁹, should be also investigated.

Another potent podosome inducer is tropoelastin⁴⁶, the soluble precursor of the cross-linked ECM protein elastin (Eln) whose expression was found accordingly downregulated in podosome deficient

$Tks5^{+/-}$ LFs (Supplementary Data 1). Moreover, and beyond individual fibrosis-modulating factors, the stiff fibrotic post-BLM aECM was shown to stimulate *Tks5* expression and podosome formation in LFs and to perpetuate the increased expression of Col1a1 (Fig. 6). In agreement, increased substrate rigidity, modeled with gelatin or polyacrylamide, has been previously shown to promote invadopodia activity⁵⁰. Therefore, the formation of podosomes in LFs upon mechanical cues from the stiff ECM of fibrotic lungs is a major component of the suggested crosstalk of ECM with fibroblasts^{51,52}, especially considering the age-related increase of ECM stiffness in the lungs⁷, and the suggested role of mechanosensitive signaling in LF activation and pulmonary fibrosis⁵³.



Given the necessity of TKS5 expression for embryonic development, the results presented here suggest that the formation of podosomes in LFs is a developmental program that gets aberrantly re-activated in IPF from pro-fibrotic factors, perpetuating LF activation and stimulating ECM invasion and LF accumulation. The re-activation of developmental pathways is a common theme in organ

fibrosis, e.g. wnt and ATX/LPA signaling in pulmonary fibrosis, providing increased plasticity and regeneration potential, as well as increased cell proliferation, migration and invasion.

Ubiquitous genetic *Tks5* haploinsufficiency was shown to attenuate BLM-induced pulmonary fibrosis with a plethora of readout assays (Fig. 4), and *Tks5*^{+/-} LFs were shown to form fewer podosomes,

Fig. 7 | Src-inhibition potently reduces podosome formation, extracellular matrix (ECM) invasion and attenuates pulmonary fibrosis. Serum starved, primary normal human lung fibroblasts (NHLFs) were pretreated for 1 h with A-419259 (SRC inhibitor) and Nintedanib, and then stimulated with recombinant human TGF β (10 ng/ml) for 24 h. **a** Graphical representation of connectivity (CMap) analysis using LINCS1000 resource of the TGF β -induced *TKSS*^{-/-} expression profile. **b, c** *TKSS* and *COL1A1* mRNA expression was interrogated with Q-RT-PCR. Values were normalized to the expression values of the housekeeping gene *B2M* and presented as fold change over control; $n = 3/3/3/3$. Statistical significance was assessed with two-tailed one-way ANOVA; $^{***}p = 0.0008$, $^{**}p = 0.0397/0.0107$ $^{*}p = 0.003$. **c** Two tailed Pearson correlation plot of *COL1A1* reduction in the same samples ($p = 0.0202$, $r = 0.84$). **d, e** *TKSS* and *COL1A1* mRNA expression was interrogated with Q-RT-PCR. Values were normalized over the expression values of the housekeeping gene *B2M* and presented as fold change over control; $n = 3/5/6/5$. Statistical significance was assessed with two-tailed one-way ANOVA; $^{*}p = 0.0202$, $^{***}p < 0.0001$. **e** Two tailed spearman correlation plot of *COL1A1* reduction in the same samples; $^{*}p = 0.0018$, $r = 0.85$. **f** Representative composite images from

double immunostaining for F-actin (red) and Cortactin (cttn; green) counter stained with DAPI (blue); arrows indicate representative podosomes; scale bars = 50 μ m. **g** Quantification of the number of podosome-containing cells per optical field; $n = 5$. Statistical significance was assessed with two-tailed one-way ANOVA; $^{***}p < 0.0001$, $^{**}p = 0.0006$, $^{*}p = 0.0014$. **h** Invasion capacity of NHLFs ($n = 4$), upon A-419259 pretreatment and TGF- β stimulation, as detected with the transwell invasion assay; $n = 4$. Statistical significance was assessed with two-tailed one-way ANOVA; $^{*}p = 0.0294$. **i** Src-inhibition attenuates pulmonary fibrosis in mouse precision cut lung slices (PCLS) generated post BLM (d11) administration. Treatment with A-419259, was administered in the first 24 h after slicing for 3 consecutive days. Representative images from PCLS stained with H&E, Fast green/Sirius red (F.G./S/R; green/red) and from double immunostaining for Tks5 and Col1a1 (green/red) are shown; scale bars=50 μ m. In all panels, representative experiment out of 2 successful independent ones are shown. In all panels, all samples are biologically independent; boxplots visualize the median of each distribution; upper/lower hinges represent 1st/3rd quartiles; whiskers extend no further than 1.5 IQR from the respective hinge. Source data for all panels are provided as a Source Data file.

resulting in diminished aECM invasion (Fig. 5), thus establishing a major pathogenetic role for Tks5-enabled podosomes and LF ECM invasion in pulmonary fibrosis. ECM invasion by non-leukocytes, a hallmark of cancer¹², is gaining increased attention in pulmonary fibrosis. IPF HLFs were shown to invade matrigel more efficiently than NHLFs or HLFs from other interstitial diseases^{8,10,11}. Enhanced invasion correlated with increased actin stress fibers¹⁰, and was suggested to be mediated, in part, by fibronectin and integrin α 4 β 1 signaling¹¹, or hyaluronan (HA) and CD44 signaling⁸. CD44 has been localized in invadopodia in breast cancer cells and has been shown to be required for invadopodia activity⁵⁴, while *TKSS*⁺ LFs were found to preferentially express CD44 (Fig. 1f), suggesting that HA/CD44 participate in the regulation of podosome formation in LFs. *TKSS*⁺ LFs were also found to express preferentially CD274/PD-L1 (Fig. 1f), a proposed marker of invasive IPF LFs⁵⁵, suggesting yet another potential signaling input for podosome formation. Moreover, BALFs from BLM-treated mice or IPF patients stimulated ECM invasion of LFs^{56,57}, shown to be attenuated upon silencing LPAR1, EGFR and FGFR2 receptors⁵⁶, or by interfering with Sdc4-CXCL10 interactions⁵⁷, suggesting additional signals that could modulate LF invasion. HER2/EGFR2, a therapeutic target in breast cancer, has been also proposed to drive invasion in LFs⁵⁸, suggesting again similarities with metastatic ADC, and further suggesting repurposing a-HER-2 agents for the treatment of IPF. Interestingly, several of the identified invasion-associated genes (Fstl3, Il11, Hbegf, Ccn2, Inhba, Podxl, Sema7a, Bcl2a1b, Bcl2a1d, Sh3rf1) were found down regulated in the *Tks5*^{-/-} invasion-defective LFs (Supplementary Data 1 and Supplementary Fig. 10d), further supporting the functional results on the role of *TKSS* and podosomes in LF ECM invasion.

Given the expression of *TKSS* in other cell types in IPF, like epithelial and basal cells (Fig. 1 and Supplementary Fig. 1) that are both intricately linked with IPF pathogenesis, a role for *TKSS* in these cell types cannot be excluded. Podosome-like structures have been reported in epithelial cells, suggested to regulate their basement membrane adhesion^{59,60}, and thus, likely, re-epithelization, an essential process in wound healing. Moreover, pharmacologic inhibition of src-kinase, a master regulator of podosomes, was shown to attenuate IPF-basal cells-induced pulmonary fibrosis in minimally BLM-injured immunodeficient mice⁶¹. However, conditional epithelial deletion of *Tks5* in future studies will be further required to dissect a possible pro-fibrotic role for *TKSS* and podosomes in these cell types.

MMP9, a podosome enriched MMP (Fig. 2), has been previously found to be expressed in IPF, localized on reactive alveolar epithelial cells, basal-like cells, clusters of alveolar macrophages, as well as subepithelial fibroblasts^{62,63}. Epithelial MMP9 has been suggested to partly mediate wound healing in keratinocytes via its proteolytic activity in podosomes, in association with CD44⁶⁰. Moreover, CD44-bound

MMP9 at the cell surface of cancer cells was shown to cleave latent TGF β , and thus promote its activation⁶⁴, and thus, possibly, the activation of subepithelial fibroblasts in IPF. On the other hand, profibrotic Thy1 fibroblasts with increased migration potential were also shown to express MMP9 following stimulation with TGF β ⁶⁵. As found here, post BLM LFs, that had increased *Tks5*/*Col1a1* levels (Supplementary Fig. 7i, m) and increased podosomes (Supplementary Fig. 7o, p), also expressed higher MMP9 levels (Supplementary Fig. 7n). Although genetic deletion of MMP9 in mice (and/or MMP2) had minimal effects in Ad-TGF β -induced pulmonary fibrosis⁶⁶, antibody-mediated MMP9 targeting demonstrated antifibrotic efficacy in a humanized immunodeficient model of IPF induced by IPF lung extracts, but only when fibrosis was promoted by “responder” IPF cells that had been shown to have reduced phosphorylated SMAD levels in response to a-MMP9 treatment *in vitro*⁶³. Further studies will be required to appreciate the role of MMP9 in the pathogenesis of IPF and the possible therapeutic potential of targeting MMP9.

Connectivity MAP (CMap) analysis has emerged as an invaluable tool to connect gene expression, drugs and disease states³³. CMap analysis of scRNAseq of IPF bronchial brushings suggested that src inhibition can reverse the observed pro-fibrotic transcriptional changes in IPF bronchial airway basal cells⁶¹. Moreover, CMap analysis of IPF transcriptional profiles and the nintedanib and pirfenidone corresponding transcriptional signatures, indicated src inhibition as the strongest connection⁶⁷. As shown here, CMap analysis of the TGF β -induced *Tks5*^{-/-} LFs profile identified, among established and promising others, src inhibition as a possible treatment to limit LF invasion (Fig. 7) and therefore pulmonary fibrosis. In agreement, src inhibition was shown to reduce *Tks5* levels and podosome numbers, to decrease aECM invasion and to reverse fibrosis in PCLS (Fig. 7) and to attenuate BLM-induced pulmonary fibrosis *in vivo* when administered by inhalation (Fig. 8). The aerosolized delivery, infrequent in IPF and animal models, opens new possibilities in IPF treatment, localizing treatment and avoiding systemic toxicity, as well as possibly increasing efficacy. The Src inhibitor Saracatinib, that has been previously also shown to attenuate BLM-induced pulmonary fibrosis⁶⁷, has recently entered clinical trials (NCT04598919), and if therapeutic effects are as efficient as in mice, the inhibition of podosome formation in LFs qualifies as a major part of its mode of action. Moreover, targeting kinase-mediated podosome formation, an inherent pathogenic LF property as shown here, as well as structural (*TKSS* and its protein-protein interactions) or effector (MMP 2/9/14) podosome components, are very promising therapeutic targets in pulmonary fibrosis.

Methods

All experimentation was performed according to the respective ethical regulations as outlined below.

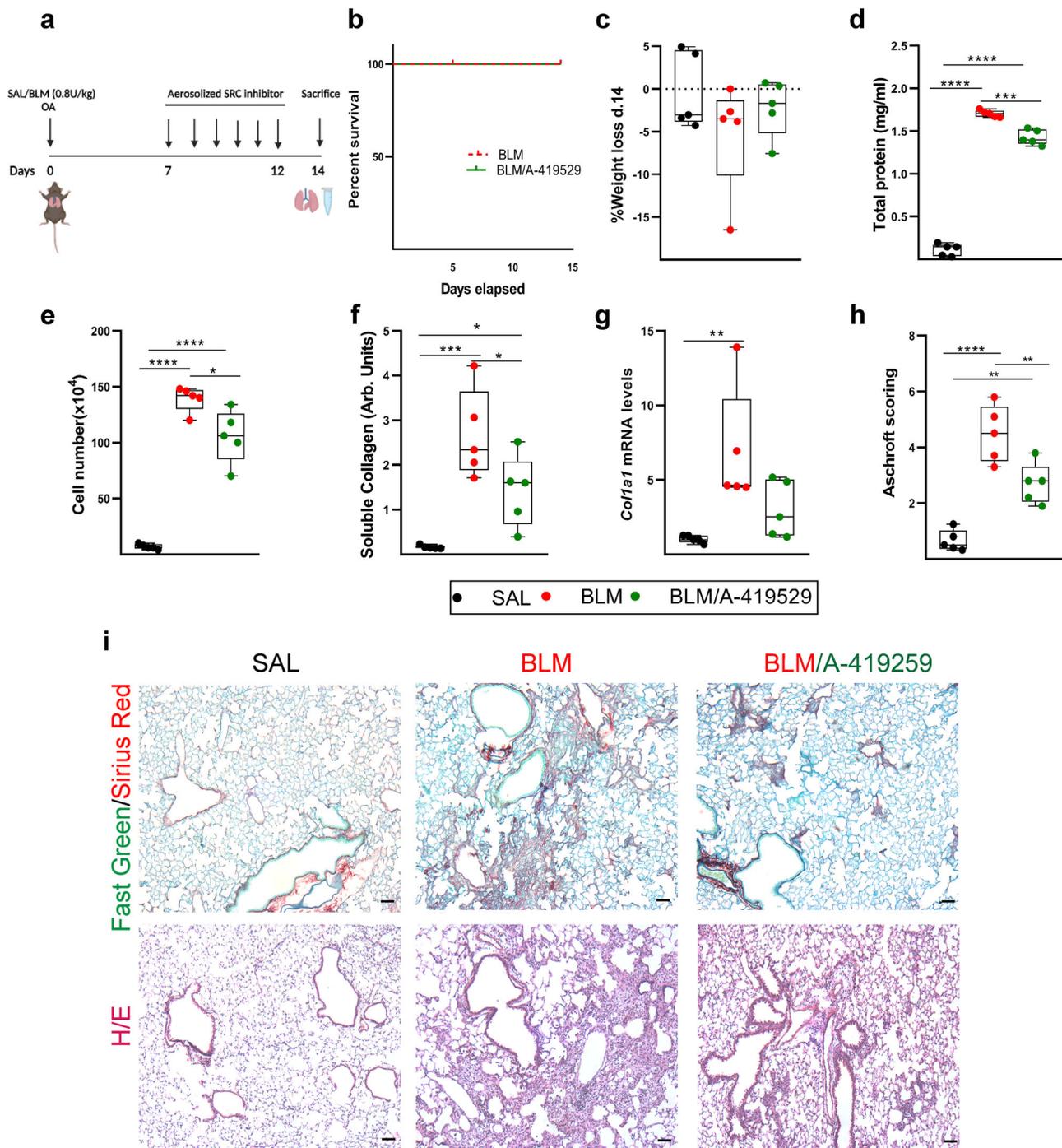


Fig. 8 | Src inhibition attenuates bleomycin (BLM)-induced pulmonary fibrosis.

a Schematic representation (biorender.com) of the BLM model and drug administration; $n = 5$. **b** Kaplan Meyer survival curve post BLM administration. **c** Weight change post BLM administration. **d** Total protein concentration in bronchoalveolar lavage fluids (BALFs), as determined with the Bradford assay. Statistical significance was assessed with two-tailed one-way ANOVA; **** $p < 0.0001$, *** $p = 0.0001$.

e Inflammatory cell numbers in BALFs, as counted with a hemacytometer. Statistical significance was assessed with two-tailed one-way ANOVA; **** $p < 0.0001$, * $p = 0.0115$. **f** Soluble collagen levels in the BALFs were detected with the direct red assay. Statistical significance was assessed with two-tailed one-way ANOVA; **** $p = 0.0004$, * $p = 0.0473$. **g** *Col1a1* mRNA expression was interrogated with Q-RT-

PCR. Values were normalized over the expression of the housekeeping gene *B2m* and presented as fold change over control. Statistical significance was assessed with two-tailed Kruskal Wallis; ** $p = 0.0094$. **h** Quantification of fibrosis severity in Hematoxylin & Eosin (H/E) stained lung sections via Ashcroft scoring. Statistical significance was assessed with two-tailed one-way ANOVA; **** $p < 0.0001$, ** $p = 0.0028/0.0074$. **i** Representative images from lung sections of murine lungs of the indicated genotypes, stained with Fast Green/Sirius Red (F.G./S.R.; green/red) and H&E; scale bars = 50 μ m. In all panels, all samples are biologically independent; boxplots visualize the median of each distribution; upper/lower hinges represent 1st/3rd quartiles; whiskers extend no further than 1.5 IQR from the respective hinge. Source data for all panels are provided as a Source Data file.

Datasets

All analyzed, re-normalized datasets (Supplementary Table 1) were sourced from Fibromine;²¹ Differentially expressed genes (DEGs): FC > 1.2, FDR corr. $p < 0.05$.

Patients

All studies with human patient samples were performed in accordance with the Helsinki Declaration principles. Lung tissue samples (Supplementary Table 2) were obtained through the University of Pittsburgh Health Sciences Tissue Bank and Yale University Pathology Tissue service, a subset of previously well characterized and published samples;⁶⁸ studies had been approved by the Yale University Institutional Review Board (Yale IRB). LFs were isolated from the lung tissue of IPF patients and from the adjacent healthy tissue of patients undergoing open lung surgery for cancer (Supplementary Table 3) at the Department of Pulmonology, Bichat-Claude Bernard Hospital, Paris/France; studies were approved by the Committee for Personal Protection (CPP)—Ile de France 1 (#0911932). All patients consented in writing to the use of their samples for research purposes; no compensation was provided for their participation.

Mice

Mice were bred at the animal facilities of Biomedical Sciences Research Center 'Alexander Fleming', under SPF conditions, at 20–22 °C, 55 ± 5% humidity, and a 12 h light-dark cycle; food (Mucedola diet #4RF21: humidity 12%, protein 18.5%, fat 3%, carbohydrate 53.5%, crude fibers 6%) and water were provided ad libitum. Mice were bred and maintained in their respective genetic backgrounds for more than 10 generations. All randomly assigned experimental groups consisted of littermate age-matched mice. The health status of the mice was monitored once per day; no unexpected deaths were observed. Euthanasia was humanly performed in a CO₂ chamber with gradual filling at predetermined time-points. All experimentation was approved by the Institutional Animal Ethical Committee (IAEC) of Biomedical Sciences Research Center "Alexander Fleming", as well as by the Veterinary Service of the governmental prefecture of Attica, Greece (# 8441/2017).

Bleomycin-induced pulmonary fibrosis

Pulmonary fibrosis was induced by a single oropharyngeal administration (OA) of 0.8 U/kg bleomycin hydrogen chloride (BLM) (Nippon Kayaku Co., Ltd., Tokyo, Japan) at day 0 into anesthetized (i.p.; xylazine, ketamine, and atropine, 10, 100, and 0.05 mg/kg, respectively) 8–10-week-old male and female mice; control groups received sodium chloride (SAL). Dose and route were selected upon prior extensive local testing to induce a solid fibrotic profile, while minimizing lethality. All randomly assigned experimental groups consisted of littermate mice. Disease development was assessed in comparison with WT littermates 14 days post-BLM, at the peak of the disease (which resolves at d21 post BLM in these settings).

The A-419259 pharmacologic src inhibitor was administered directly in the mouse lungs through inhalation, using the inExpose system (Scireq, Cat.Number IX-XNI-T6) to conscious, softly re-strained mice. The inhibitor was administered at a therapeutic mode, once daily for 6 consecutive days, starting from day 7 post BLM administration. A-419259 was diluted in saline at a concentration of 0.182 mg/ml and 4 ml was administered for 5 mins in a group of 6 mice, corresponding to a final dose of 2 mg/kg per mouse. Control groups received aerosolized saline.

Following weighing, respiratory functions were measured with FlexiVent (SCIREQ, Montreal, Canada), according to manufacturer instructions and as previously described^{24,43}. Briefly, a pressure-volume loop (PV) perturbation and a forced oscillation technique (FOT, single and low frequency) were applied in tracheotomized mice to produce the indicated measurements.

Bronchoalveolar Lavage fluid (BALF) was obtained by lavaging the lungs with 1 ml of 0.9% sterile sodium chloride three times. After the isolation, the samples were centrifuged at 1200 g for 10 min at 4 °C, the first BALF supernatant was stored at -80 °C for protein and collagen measurements. To estimate pulmonary inflammation, BALF cell pellets were redissolved in 1 ml saline, stained with 0.4% Trypan Blue solution and were counted with the use of a Neubauer hemacytometer. Total protein levels in BALFs, an indication of pulmonary edema and vascular leak, were assessed with the Bradford assay according to the manufacturer's instructions (Bio-Rad, Hercules, CA, USA). In a 96-well plate 5 µl of every BALF sample is placed, followed by the addition of 245 µl of 1x Bradford reagent (Serva/39222.03) and incubation for 5 minutes in the dark. Absorbance values were then measured at 595 nm, using a spectrometer, and were converted in concentration values (mg/ml) using a bovine serum albumin standard curve (BSA 0–2 mg/mL). Total soluble collagen in BALFs was quantified using the Sirius Red assay. 50 µl of BALF samples, diluted in 350 µl of 0.5 M acetic acid, were incubated for 30 min with 400 µl of Direct (365548-5 G Sigma-Aldrich) at RT, in the dark. This was followed by centrifugation, at 12000 g for 10 min and isolation of 200 µl of the supernatants. Absorbance values were measured at 540 nm, using a spectrometer, and were converted in concentration values (µg/ml) using a rat tail collagen I (C7661-5mg Sigma Aldrich) standard curve (0–500 µg/mL).

Histology

The right lung was fixed overnight in 10% neutral buffered formalin and embedded in paraffin. 5µm lung sections were cut using a Microtome and stained with Hematoxylin/eosin (H&E) (Papanicolaou's solution HX16967353 Sigma Aldrich/ Eosin G CI45380 ROTH) with standard protocols. Fibrosis development was quantified by two independent reviewers, in a blinded manner, based on a modified Ashcroft score (0, normal lung; 1, isolated alveolar septa with gentle fibrotic changes; 2, fibrotic changes of alveolar septa with knot-like formation; 3, contiguous fibrotic walls of alveolar septa; 4, single fibrotic masses; 5, confluent fibrotic masses; 6, large contiguous fibrotic masses; 7, air bubbles; 8, fibrous obliteration). For Fast green-Sirius red (F.G/S.R) collagen staining, lung sections were deparaffinized in xylene and ethanol and incubated in Bouin's solution (75% picric acid/ 25% formaldehyde/ 1% acetic acid), for 1 hour at 56 °C, followed by staining with Fast Green (Glenthams Life Sciences GT3407/100 g) 0.04% in picric acid for 15 minutes and Sirius Red 0.1%/Fast Green 0.04% dissolved in picric acid (197378-100 g Sigma-Aldrich) for 40 minutes. Stained sections were washed in acetic acid, then dehydrated and mounted with DPX (06522-500 ml Sigma-Aldrich). PCLS, isolated and cut as described below stained with H&E. For X-gal (Lac-Z), freshly isolated mouse lungs were inflated with 0.1 g/ml sucrose in 50% OCT/PBS, followed by the simultaneous embedding and freezing in OCT, using isopentane and dry ice. Sections of 6–10 µm were cut using a cryotome and fixed in 2% formaldehyde/ 0.2% glutaraldehyde for 15 minutes at 4 °C. Next, they were washed twice in cold PBS/ 2 mM MgCl₂ (PENTA) for 10 minutes and stained overnight with X-gal staining solution (2 mg/mL X-gal in 0.1 M Sodium phosphate buffer pH=7.3, 0.01% Sodium deoxycholate (30970 FLUKA), 5 mM K₃Fe(CN)₆ (60300 FLUKA), 5.7 mM K₄Fe(CN)₆ (12639 Riedel-de Haen), 2 mM MgCl₂, 0.02% NP-40 (UN3082 Applichem) at 37 °C in the dark. The sections were then rinsed twice with PBS/2 mM MgCl₂ and dH₂O for 5 minutes at room temperature, counterstained with eosin following by dehydration and mounting with DPX. Imaging was performed using a Nikon Eclipse E800 microscope (Nikon Corp., Shinagawa-ku, Japan) attached to a Q Imaging EXI Aqua digital camera, using the Q-Capture Pro 7 software. For immunohistochemistry studies, lung sections were deparaffinized in xylene, rehydrated in a gradient of ethanol, and briefly washed with water. The slides were kept in tap water until ready to

perform antigen retrieval with sodium citrate buffer with pH 6.0 by autoclave for 20 min. Then they were treated with blocking solution (10% normal goat serum/2% BSA; A9647 Sigma Aldrich) at room temperature for 1 h and incubated with primary antibodies overnight at 4 °C. After washing, they were incubated with fluorophore-conjugated secondary antibodies diluted in the blocking solution. Following this, sections were washed 3 times with PBS-T and mounted with medium containing DAPI (F6057, Sigma Aldrich) for nuclear visualization. Imaging was performed using a TCS SP8X White Light Laser confocal system (Leica).

In vitro/ex vivo lung fibroblast cell model

Normal human lung fibroblasts (NHLFs) and IPF-HLFs were isolated from fresh tissue samples by plating several 2-3 mm pieces on 10 cm tissue culture plates in DMEM supplemented by penicillin/streptomycin solution, 10% FBS at 37 °C and 5% CO₂ in a humidified atmosphere. 10-14 days following plating, proliferating fibroblasts surrounded the tissue pieces, which were then removed, and cells are detached with trypsin-EDTA solution and replated in F75 tissue culture flasks (PO) until confluent. Removed tissue pieces were replated in fresh 10 cm tissue culture plates for a second round of fibroblast outgrowth in the same conditions as above. Following 2-3 passages, homogeneous fibroblasts' colonies are observed that can be frozen down; upon thawing 3-4 passages are needed to obtain adequate numbers for experimentation; HLFs are used until passage 7-8. A similar procedure was independently used for an additional NHLF clone (Fig. S3A-B), as previously published⁶⁹.

Primary normal mouse lung fibroblasts (NMLFs) were isolated from 8-10-week-old C57Bl6/J mice and/or from BLM-challenged mice. Upon sacrifice, perfused lungs were excised in DMEM. Then lungs were minced and digested with 0.7 mg/ml collagenase type IV (C5138-Sigma), for 1 h at 37 °C. Digestion was followed by filtration and the suspension was centrifuged at 1200 g for 5 min. Finally, the pellet was resuspended in DMEM (GIBCO 41966-029) containing 10% FBS (GIBCO 10437-028/ origin Mexico). All experiments were performed at passage 2-3.

Murine and human embryonic LFs cell lines, 3T3 and MRC5 respectively, were purchased from ATCC (#CRL-1658 and #CCL-171 respectively).

NHLFs, IPF-HLFs, NMLFs, 3T3 and MRC5 cells were cultured in DMEM supplemented with 10% fetal bovine serum (FBS) and streptomycin/penicillin (GIBCO 15140-122) and amphotericin (GIBCO 15290-018) and incubated at 37 °C and 5% CO₂. Cells were cultured to 60-80% confluency, were starved overnight with serum-free DMEM (+ 0.1% BSA) and were exposed to 10 ng/ml recombinant human TGF-β1 (240-B-002 R&D SYSTEMS) for 24 h in serum-free DMEM. In control samples the diluent of TGFβ (7.5% BSA in H₂O) was used. For pharmacologic studies, NHLFs were seeded at 6-well plates, were serum starved and pre-treated for 1 h with the indicated increasing concentrations of various agents and their diluents in controls. After one hour of pre-treatment, cells were incubated with TGFβ as usual.

The proliferation of all LF cultures was quantified with the MTT assay in 96-well plates, where a common solution of OPTI-MEM (GIBCO 11058-021) and MTT (0.7 mg/ml) (ACROS ORGANICS158990010) was added into each well. After incubation for 4 h and having confirmed the formation of purple crystals, the media was removed, and acidified isopropanol was added into each well to dissolve the formazan crystalline product. Absorbance values were determined at 570 nm and background subtracted at 660 nm using an OPTImax (Microplate Photometer (Molecular Devices).

The spreading and migration capabilities of LFs were assessed in a scratch wound assay. Cells were cultured into a 12-well plate as above and left to grow. Upon confluency, TGFβ was added, as described above. A wound was generated using a sterile pipette tip. The plate was

placed in a tissue culture incubator at 37 °C, and photos were taken under a reverse microscope at specific time intervals.

Migration and invasion were quantified also with Boyden chambers (Costar 20122024/REF3422) according to manufacturer's instructions, as shown schematically in Fig. 5E. Briefly, LFs were added to the upper chamber which was pre-coated with aECM substrate (or not for migration) and allowed for 6 hours to invade/migrate through the transwell membrane to the lower side which was in touch with the starvation medium. Then, the cells which remained in the upper chamber were removed, while invasive or migratory cells, after washes and fixation, were stained with crystal violet. Additionally, stained cells were lysed with Lysis Buffer for 20 minutes and absorbance was measured at 550 nm using the TECAN Sunrise Microplate Photometer.

The proteolytic capacity of LFs was assessed with the Fluorescein gelatin degradation assay. 12-mm glass coverslips were acid washed with 20% nitric acid, incubated with 50 µg/ml poly-L-lysine, crosslinked with 0.5% glutaraldehyde and then coated for 20 min with 0.2% fluorescein-conjugated gelatin (Invitrogen, Gelatin from Pig Skin, Fluorescein Conjugate: G13187) in 2% sucrose (131621.1211Panreac) -containing phosphate buffered saline (PBS). Then, the coverslips were treated with sodium borohydride (NaBH₄-Sigma 452882), washed with PBS, and transferred to a new 24-well-plate. Fibroblasts were seeded in gelatin-coated coverslips and treated with TGF-β1 as previously described. After 24 hours, medium was replaced with full medium and cells were processed for immunofluorescence, 24-48 hours later. The quantification of the degraded gelatin was analyzed using ImageJ, as previously described⁷⁰. Briefly, the measurements of the degraded area, are reflected from the measurement of the "area fraction", which has been followed from the threshold adjustment that represents the actual degradation. Following this process, the "area fraction" values were then normalized to the number of nuclei as measured from the Dapi channel of the correlated image.

For immunofluorescence staining, cells were seeded in coverslips (20.000 cells/well) and after starvation, incubated with TGF-β1 as usual. The cells were fixed with 4% PFA for 15 min and permeabilized with 0.1% Triton X (T8532, Sigma Aldrich) for 10 min. This was followed by blocking with 2% BSA in PBS for 1 h at RT. The cells were incubated overnight at 4 °C with primary antibodies. The next day, cells were washed, and incubated with a secondary antibody and conjugated phalloidin in 1% BSA/PBS for 60 min at RT. Finally, after washing, coverslips were mounted with a drop of mounting Fluoroshield medium (containing DAPI for nucleus labeling). Confocal microscope images were analyzed with ImageJ. Signal colocalization was performed via the orthogonal views and k-curves analysis with ImageJ tools showing the intensity per z-stacking or per distance of the maximal projection accordingly.

The decellularization and generation of acellular ECM (aECM) from mouse lungs was performed based on similar protocols for other tissues^{71,72}. Briefly, whole lungs were isolated and treated with increasing concentrations of SDS (Fisher Bioreagents BP166) (0.01, 0.1, 1%) in a PBS solution, with 24 h incubation for each SDS concentration. For the final step, decellularized lung tissues were washed with PBS for at least 3 days, cut into small pieces and stored at -80 °C. Frozen tissue was lyophilized using a lyophilizer and then milled in liquid nitrogen. To produce an ECM substrate, the milled form of the matrix was solubilized through enzymatic digestion. Pepsin (Sigma-Aldrich, P6887) was dissolved in 0.1 M HCl to make a concentration of 1 mg/ml. Approximately 10 mg of the ECM powder were digested in 1 mL of pepsin solution, in order to solubilize the ECM components. After approximately 48 hours, the matrix was diluted using 0.1 M acetic acid to make a 5 mg/ml concentration of lung ECM solution, which was used as a coating substrate for cells.

Precision cut lung slices

C57-Bl/6, 8-10-week-old mice, were administered with SAL/BLM as described above. On day 11, mice were sacrificed and lungs, after perfusion, were inflated with 1 ml of with 1.5% Low Melting Agarose (15517-014-Invitrogen) in saline. Lungs were then isolated and incubated at RPMI medium (Thermo Fisher Scientific) supplemented with 10% FBS and 1% penicillin-streptomycin for 30 min at 4 °C to allow agarose polymerization. The left lobe of the agarose filled lungs, was cut into 200 µm slices (PCLS) with Vibratome. PCLS were then cultured in 700 µL of RPMI (GIBCO-21875-034) medium supplemented with 10% FBS and 1% penicillin-streptomycin in 24-well plates at standard conditions (37 °C and 5% CO₂) overnight. Next, PCLS were incubated to 2 µM of A-419259 (Src-family inhibitor) and H₂O for 3 consecutive days, changing the treatment daily. In day 14, PCLS were fixed overnight with PFA (Sigma-Aldrich P6148) at 4 °C, until slices embedded in paraffin. Finally, 5 µm sections of PCLS were cut using a Microtome and stained for H&E, F.G/S.R and antibodies accordingly.

Antibodies- reagents

Antibodies used in this study included anti-Tks5 (SH3 domain) rabbit monoclonal antibody (Merck, 3174822, 1:100), anti-SH3PXD2A mouse monoclonal antibody (Origene, clone OTIIF5-TA811757S, LOT F001, 1:250), colla1 rabbit polyclonal antibody (Invitrogen, PA5-29569, LOT XK3738717 1:100), anti-A-actin (sma) mouse monoclonal antibody, (Origene, clone UM870129, LOT F001 1:250), recombinant Anti-Cortactin rabbit monoclonal (EPI922Y) antibody (Abcam, ab81208, 1:500), Alexa Fluor™ 633 Phalloidin (Invitrogen, A22284, LOT 2274768 1:50), MMP-9 XP Rabbit monoclonal (D6O3H) Antibody (Cell signaling #13667, LOT 3 1:100). Secondary antibodies included: Goat anti-Rabbit IgG (H + L) Cross-Adsorbed Secondary Antibody Alexa Fluor 488 (Life technologies, A11008, LOT1470706), Goat anti-Rabbit IgG (H + L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 555 (Life technologies, A21428, LOT 1670185), Goat anti-Mouse IgG (H + L) Highly Cross-Adsorbed Secondary Antibody Alexa Fluor 488 (Life technologies, A11029, LOT1705900) Goat anti-Mouse IgG (H + L) Highly Cross-Adsorbed Secondary Antibody Alexa Fluor 555 (Life technologies, A22424, LOT1726548); all used in 1:500 dilution; all antibodies were selected from previous publication, while all antibodies were validated by the corresponding manufacturers for the specific employed methodologies (IHC/IF). For pharmacological studies A-419259 inhibitor SML0446 (Sigma-Aldrich), Nintedanib SML2848 (Sigma-Aldrich) and Pirferidone P2116 (Sigma-Aldrich) were used.

Real Time quantitative RNA RT-PCR

In human samples, RNA was extracted from 30 to 50 mg of frozen lung tissue in 700 µL of Qiazol (Lysis buffer, Qiagen, Valencia, CA) by tissue disruption and homogenization using an electric homogenizer (Poly-Tron homogenizer H3660-2A, Cardinal Health, Dublin, OH) at 15.000 g for 15 seconds, according to the manufacturer's instructions. RNA was purified using the miRNeasy Mini kit (217004, Qiagen, Valencia, CA) with the assistance of the Qiacube automated system (9001292, Qiagen, Valencia, CA). The purity of the RNA was verified using NanoDrop at 260 nm and the quality of the RNA was assessed using the Agilent 2100 Bioanalyzer (Agilent, Technologies, Santa Clara, CA). Real-time PCR was performed with Taqman primers as described in the table below. Values were normalized to the expression of B2M.

In mouse samples RNA was extracted from the left lung lobe using the Tri Reagent (TR-118) obtained from Invitrogen and treated with DNase (RQ1 RNase-free DNase) prior to RT-PCR according to manufacturer's instructions. cDNA synthesis was performed using 2 µg of total RNA per sample in 20-µl reaction using M-MLV RT (Promega). Real-time PCR was performed on a BioRad CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad Laboratories). Values were normalized to the expression of b-2 microglobulin (b2m). The annealing temperature for all primers was 58 °C. Primers

sequences for RT and genomic PCRs are depicted at Supplementary Table 5.

RNA sequencing

Six total RNA samples were prepared, and their concentration was measured with nanodrop (ND1000 Spectrophotometer-PEQLAB). The samples measured to a concentration of 400-500 ng/µl and therefore 1 µl of RNA, from each sample was used to proceed with the library preparation. The RNA quality of each sample was measured in bioanalyzer (Agilent Technologies) using the Agilent RNA 6000 Nano Kit reagents and protocol. For the preparation of per sample libraries, the 3' mRNA-Seq Library Prep Kit Protocol for Ion Torrent (QuantSeq-LEXOGEN™ Vienna, Austria) was used according to manufacturer's instruction. Briefly, library generation was initiated by oligodT priming which contains the Ion Torrent compatible linker sequences. 5 to 500 ng per 5 µl of RNA from each sample was used to perform the first strand synthesis. After first strand synthesis any remaining RNA was removed and second strand synthesis was initiated by a random primer, containing Ion Torrent compatible linker sequences at its 5' end, and a sequence polymerase. In line barcodes were introduced at this point. Second strand synthesis was followed by a magnetic bead-based purification step and the resulted purified library was amplified for 14 cycles and re-purified. Quality and quantity of each library was assessed in a bioanalyzer using the DNA High Sensitivity Kit reagents and protocol (Agilent Technologies). The quantified libraries were pooled together at a final concentration of 7 pM. The libraries pool was processed on the Ion Proton One Touch system where the libraries were templated and enriched using either the Ion PI™ Hi-Q™ OT2 200 Kit (ThermoFisher Scientific) and sequenced, with the Ion PI™ Hi-Q™ Sequencing 200 Kit on Ion Proton PI™ V2 chips (ThermoFisher Scientific) according to commercially available protocols. 3' RNA-sequencing was performed on an Ion Proton™ System⁷³, according to the manufacturer's instructions. Initial analysis took place in Ion Torrent server.

Single cell RNA-seq data re-analysis

Single cell RNA-seq data were downloaded from [GSE122960](#) and processed with the R package Seurat (v.3.1.2 & 4.0.5)^{74,75}. A similar to the original data analysis strategy was applied. Initially, each sample was processed on its own. After removing low-quality cells and genes, data were normalized using the LogNormalize method of Seurat and then top variable features were selected using the vst method. Data were scaled prior to principal component analysis (PCA) application and selection of the top principal components. The latter were used for a Shared Nearest Neighbor (SNN) graph-guided cell clustering and last, t-SNE dimensionality reduction was performed. Cell typing followed that of the original analysis as much as possible. Samples integration was performed with the standard Seurat v3 integration pipeline first for samples within each and then across phenotypes (donor and IPF). Integrated data were re-scaled, clustering and dimensionality reduction were repeated. Cell types inherited from single-sample analysis were validated and corrected whenever required according to the original publication of the dataset. Marker genes were identified using the Wilcoxon Rank Sum test applied on the "RNA" slot of the integrated samples object. Absolute fold change of at least 1.2 on natural scale and Bonferroni-corrected *p*-value < 0.05 were used as differential expression thresholds. Fibroblast sub-clusters were identified with a resolution of 0.1 after fibroblast cells were isolated from the rest of the dataset, re-scaled and new variable features were found as above.

Data processing

Quant-Seq (Lexogen) FASTQ files obtained from the Ion Proton sequencing procedure were trimmed with Trim Galore (v.0.6.51) to remove low quality read ends using a Phred score of 20. Subsequently, a two steps alignment procedure was applied. Pre-processed reads

were aligned against the GRCm38 reference genome (Ensembl) with HISAT2 (v.2.2.1)⁷⁶ and then the reads left unmapped were subjected to a second alignment round using BOWTIE2 (v.2.3.5.1)⁷⁷ with the `-local` and `-very-sensitive local` switches turned on.

Computational analysis

Downstream analysis of the resulting BAM files was performed with metaseqR2 (v.1.9.2)⁷⁸. Briefly, the raw BAM files, one per sample, were summarized to a 3' UTR reads count table, using the package `GenomicRanges` (v.1.44.0)⁷⁹ and Ensembl mouse genome mm10. For the UTR counting, the entire 3' UTR region, with a minimum length of 300 base pairs and 50 base pairs to flank the UTR end, was taken into consideration. In the resulting reads count table, each row represented one 3' UTR region and each column a Quant-Seq sample. Next, reads were summarized per gene and the returned gene count table was normalized using the package `EDA-Seq` (v.2.26.1)⁸⁰ after removing genes having zero reads across all samples. Post-normalization, gene counts were filtered for possible artifacts using default gene filtering options. The filtered gene counts table was subjected to differential expression analysis using sequentially all nine individual statistical analysis methods supported by metaseqR2. Their p-values were then combined by the PANDORA algorithm to account, among others, for the false positives reported. Benjamini-Hochberg corrected PANDORA p-values of less than 0.05 and absolute fold change of at least 1.2 were used as differential expression thresholds. Normalized expression values required for each heatmap were retrieved and standardized across samples. Hierarchical clustering of samples and genes based on calculated Euclidean distance.

Gene set enrichment analysis

Differential expression analysis results were sorted by decreasing fold change and used for Gene Set Enrichment Analysis (GSEA) against Gene Ontology terms⁸¹ using the package `clusterProfiler` (v.4.0.5)⁸². Signed normalized enrichment score (NES) was used to isolate the top of the significantly enriched induced (NES > 0) and suppressed (NES < 0) terms (adjusted p-value < 0.05).

Text mining

PubMed 2022 baseline was downloaded from the respective FTP site. XML R package (v.3.99.0.8) was used to create an abstract-based corpus which was then queried with `rentrez` R package (v.1.2.3) for *IPF[All Fields] OR ("pulmonary fibrosis"[MeSH Terms] OR "pulmonary fibrosis"[All Fields]) OR ("lung diseases, interstitial"[MeSH Terms] OR "interstitial lung diseases"[All Fields] OR "interstitial lung disease"[All Fields])* containing elements. Subsequently, human HGNC gene symbols atomization was performed using `pubmed.mineR` package⁸³ (v.1.0.19) and recovered genes were intersected with the human homologs of the mouse Quant-seq differentially expressed genes. Homolog feature mapping was obtained via `biomaRt` R package⁸⁴ (v. 2.48.3).

Transcription factor analysis

Transcription factor analysis was performed using DoRothEA R package⁸⁵ (v.1.4.2). All mouse transcription factor-target regulons were queried for those high quality ones ("A" level of confidence) including as targets any of the Quant-seq differentially expressed genes. Subsequent filtering maintained those interaction pairs where both interactors were found significantly deregulated in the bulk sequencing experiment. Based on the mode of regulation (mor) as described in the DoRothEA database and the Quant-seq derived differential expression data, pairs with the same or opposite direction of deregulation were maintained in cases of an activator or repressor transcription factor, respectively.

CMap/LINCS analysis

CMap/LINCS database (<https://clue.io/query>) query was performed using the top 150 up and top 150 down regulated genes as sorted by fold change. 150 consisted of the maximum number of genes supported by the platform. L1000 gene expression data from the Expanded CMap LINCS Resource 2020 (last update 11/23/2021) were queried. Results were processed by the R package `cmapR` (v.1.4.0). FDR-corrected p-values less than 0.05 were used to isolate those signatures having a statistically important connection with the provided one and signed normalized connectivity score (NCS) was used to discern between similar and opposite signatures. Focus was given on compounds (`trt_cp`) and peptides (`trt_lig`) having an already known mechanism of action and known targets. Datasets comprising of only one replicate and/or of a treatment duration other than 24 hours were discarded.

Statistics

Statistical significance was assessed with the Prism (GraphPad) software according to its built-in recommendations, as detailed at each figure legend. Briefly, and unless otherwise stated, all datasets were tested for normal distributions via Shapiro-Wilk test, while all measurements were performed on distinct and independent samples. For the comparison of two normally distributed experimental groups, we employed the two-tailed unpaired t-test, for equal SDs or Welch's t-test for unequal ones. Not normally distributed data were analyzed with the two-sided Mann-Whitney test. Normally distributed multi parametric data with equal SDs were analyzed with the unpaired one-way ANOVA test and post hoc Tukey's test. Welch's ANOVA test coupled with post hoc Games-Howell test was used in cases of unequal SDs. For non-normally distributed multiparametric data, Kruskal-Wallis, followed by post hoc Dunn's test was utilized. For correlation analysis we used Pearson's, for normally distributed, or Spearman's correlation for non-normally distributed datasets. Most data are presented on box and whiskers graphs depicting the median as well as all experimental values (n).

Image creation

Third party images were created at bioRender.com (under the relative agreements; 31/08/2023): Fig. 4a (DH25SQBJQT), Fig. 5e (XG25SQ37G), Fig. 8a (UC25SQ61BR), Sup Fig. 8a (VP25SQ8NTX), Sup Fig. 8b (GM25SQ96G1), Sup Fig. 8d (AV25SQA9RX), Sup Fig. 8f (Z125SQ9VL2), and Sup Fig. 9a (BQ25SQ7828).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All re-analyzed publicly available datasets are listed in Supplementary Table 1, including accession numbers and hyperlinks. Quant-Seq data have been deposited at the GEO database under the accession code [GSE220982](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE220982). Already published, re-analyzed single cell RNA-seq data used are available at GEO database: [GSE122960](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE122960). All other relevant experimental data are within the paper and its supplementary information files. Source data are provided with this paper.

Code availability

Data and code for the recreation of the computationally-created figures of the paper have been deposited at Zenodo (<https://doi.org/10.5281/zenodo.8296510>)⁸⁶ and at https://github.com/dfanidis/TKS5_podosomes_IPF. Detailed scripts used are unrestrictedly available, within 10 days, upon request to Dionysios Fanidis (fanidis@fleming.gr); Institute for Fundamental Biomedical Research, Biomedical Sciences Research Center Alexander Fleming, Athens, Greece).

References

1. Rockey, D. C., Bell, P. D. & Hill, J. A. Fibrosis — a common pathway to organ injury and failure. *N. Engl. J. Med.* **372**, 1138–1149 (2015).
2. Raghu, G. et al. Diagnosis of idiopathic pulmonary fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am. J. Respir. Crit. Care Med.* **198**, e44–e68 (2018).
3. Martinez, F. J. et al. Idiopathic pulmonary fibrosis. *Nat. Rev. Dis. Prim.* **3**, 17074 (2017).
4. Plikus, M. V. et al. Fibroblasts: Origins, definitions, and functions in health and disease. *Cell* **184**, 3852–3872 (2021).
5. Tomasek, J. J., Gabbiani, G., Hinz, B., Chaponnier, C. & Brown, R. A. Myofibroblasts and mechano-regulation of connective tissue remodelling. *Nat. Rev. Mol. Cell Biol.* **3**, 349–363 (2002).
6. Penke, L. R. & Peters-Golden, M. Molecular determinants of mesenchymal cell activation in fibroproliferative diseases. *Cell Mol Life Sci* **76**, 4179–4201 (2019).
7. Pardo, A. & Selman, M. Lung fibroblasts, aging, and idiopathic pulmonary fibrosis. *Ann. Am. Thorac. Soc.* **13**, S417–s421 (2016).
8. Li, Y. et al. Severe lung fibrosis requires an invasive fibroblast phenotype regulated by hyaluronan and CD44. *J. Exp. Med.* **208**, 1459–1471 (2011).
9. Lovgren, A. K. et al. β -arrestin deficiency protects against pulmonary fibrosis in mice and prevents fibroblast invasion of extracellular matrix. *Sci. Transl. Med.* **3**, 74ra23 (2011).
10. Karvonen, H. M. et al. Myofibroblasts in interstitial lung diseases show diverse electron microscopic and invasive features. *Lab. Invest.* **92**, 1270–1284 (2012).
11. White, E. S. et al. Integrin α 4 β 1 regulates migration across basement membranes by lung fibroblasts: a role for phosphatase and tensin homologue deleted on chromosome 10. *Am. J. Respir. Crit. Care Med.* **168**, 436–442 (2003).
12. Hanahan, D. Hallmarks of cancer: new dimensions. *Cancer Discov* **12**, 31–46 (2022).
13. Karampitsakos, T. et al. Lung cancer in patients with idiopathic pulmonary fibrosis. *Pulm. Pharmacol. Ther.* **45**, 1–10 (2017).
14. Murphy, D. A. & Courtneidge, S. A. The ‘ins’ and ‘outs’ of podosomes and invadopodia: characteristics, formation and function. *Nat. Rev. Mol. Cell Biol.* **12**, 413–426 (2011).
15. Paterson, E. K. & Courtneidge, S. A. Invadosomes are coming: new insights into function and disease relevance. *FEBS J.* **285**, 8–27 (2018).
16. Hoshino, D., Branch, K. M. & Weaver, A. M. Signaling inputs to invadopodia and podosomes. *J. Cell. Sci.* **126**, 2979–2989 (2013).
17. Di Martino, J. et al. The microenvironment controls invadosome plasticity. *J. Cell. Sci.* **129**, 1759–1768 (2016).
18. Murphy, D. A. et al. A Src-Tks5 pathway is required for neural crest cell migration during embryonic development. *PLoS ONE* **6**, e22499 (2011).
19. Cejudo-Martin, P. et al. Genetic disruption of the sh3pxd2a gene reveals an essential role in mouse development and the existence of a novel isoform of tks5. *PLoS ONE* **9**, e107674 (2014).
20. Li, C. M. et al. Differential Tks5 isoform expression contributes to metastatic invasion of lung adenocarcinoma. *Genes Dev.* **27**, 1557–1567 (2013).
21. Fanidis, D., Moulos, P. & Aidinis, V. Fibromine is a multi-omics database and mining tool for target discovery in pulmonary fibrosis. *Sci. Rep.* **11**, 21712 (2021).
22. Reyfman, P. A. et al. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 1517–1536 (2019).
23. Mouratis, M. A. & Aidinis, V. Modeling pulmonary fibrosis with bleomycin. *Curr. Opin. Pulm. Med.* **17**, 355–361 (2011).
24. Barbayianni, I., Ninou, I., Tzouvelekis, A. & Aidinis, V. Bleomycin revisited: a direct comparison of the intratracheal micro-spraying and the oropharyngeal aspiration routes of bleomycin administration in mice. *Front. Med.* **5**, 269 (2018).
25. Tashiro, J. et al. Exploring animal models that resemble idiopathic pulmonary fibrosis. *Front. Med.* **4**, 118 (2017).
26. Seals, D. F. et al. The adaptor protein Tks5/Fish is required for podosome formation and function, and for the protease-driven invasion of cancer cells. *Cancer Cell* **7**, 155–165 (2005).
27. Auwerx, J. et al. The European dimension for the mouse genome mutagenesis program. *Nat. Genet.* **36**, 925–927 (2004).
28. Raess, M., de Castro, A. A., Gailus-Durner, V., Fessele, S. & Hrabě de Angelis, M. INFRAFRONTIER: a European resource for studying the functional basis of human disease. *Mamm. Genome* **27**, 445–450 (2016).
29. Ehlich, H. et al. INFRAFRONTIER quality principles in systemic phenotyping. *Mamm. Genome* **33**, 120–122 (2022).
30. Schwenk, F., Baron, U. & Rajewsky, K. A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* **23**, 5080–5081 (1995).
31. Ryder, E. et al. Rapid conversion of EUCOMM/KOMP-CSD alleles in mouse embryos using a cell-permeable Cre recombinase. *Transgenic Res.* **23**, 177–185 (2014).
32. Farley, F. W., Soriano, P., Steffen, L. S. & Dymecki, S. M. Widespread recombinase expression using FLP_{er} (flipper) mice. *Genesis* **28**, 106–110 (2000).
33. Subramanian, A. et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452.e1417 (2017).
34. Wells, A. U. et al. Nintedanib in patients with progressive fibrosing interstitial lung diseases-subgroup analyses by interstitial lung disease diagnosis in the INBUILD trial: a randomised, double-blind, placebo-controlled, parallel-group trial. *Lancet Respir. Med.* **8**, 453–460 (2020).
35. Li, H. et al. Src family kinases and pulmonary fibrosis: a review. *Biomed. Pharmacother.* **127**, 110183 (2020).
36. Foxall, E. et al. PAK4 kinase activity plays a crucial role in the podosome ring of myeloid cells. *Cell Rep.* **29**, 3385–3393.e3386 (2019).
37. Rafiq, N. B. M. et al. Podosome assembly is controlled by the GTPase ARF1 and its nucleotide exchange factor ARNO. *J. Cell Biol.* **216**, 181–197 (2016).
38. Rottiers, P. et al. TGF β -induced endothelial podosomes mediate basement membrane collagen degradation in arterial vessels. *J. Cell. Sci.* **122**, 4311–4318 (2009).
39. Varon, C. et al. Transforming growth factor β induces rosettes of podosomes in primary aortic endothelial cells. *Mol. Cell Biol.* **26**, 3582–3594 (2006).
40. Charbonneau, M. et al. Platelet-derived growth factor receptor activation promotes the prodestructive invadosome-forming phenotype of synoviocytes from patients with rheumatoid arthritis. *J. Immunol.* **196**, 3264–3275 (2016).
41. Quintavalle, M., Elia, L., Condorelli, G. & Courtneidge, S. A. Micro-RNA control of podosome formation in vascular smooth muscle cells in vivo and in vitro. *J. Cell. Biol.* **189**, 13–22 (2010).
42. Spuul, P. et al. VEGF-A/notch-induced podosomes proteolyse basement membrane collagen-IV during retinal sprouting angiogenesis. *Cell Rep.* **17**, 484–500 (2016).
43. Zannikou, M. et al. MAP3K8 regulates Cox-2-mediated prostaglandin E(2) production in the lung and suppresses pulmonary inflammation and fibrosis. *J. Immunol.* **206**, 607–620 (2021).
44. van Helden, S. F. G. et al. A critical role for prostaglandin E2 in podosome dissolution and induction of high-speed migration during dendritic cell maturation. *J. Immunol.* **177**, 1567–1574 (2006).
45. Lebel, M. et al. Invadosome formation by lung fibroblasts in idiopathic pulmonary fibrosis. *Int. J. Mol. Sci.* **24**, 499 (2022).

46. Iizuka, S. et al. Crosstalk between invadopodia and the extracellular matrix. *Eur. J. Cell. Biol.* **99**, 151122 (2020).
47. Artym, V. V. et al. Dense fibrillar collagen is a potent inducer of invadopodia via a specific signaling network. *J. Cell. Biol.* **208**, 331–350 (2015).
48. Juin, A. et al. Discoidin domain receptor 1 controls linear invadopodia formation via a Cdc42-Tuba pathway. *J. Cell. Biol.* **207**, 517–533 (2014).
49. Denny, W. A. & Flanagan, J. U. Inhibitors of Discoidin Domain Receptor (DDR) kinases for cancer and inflammation. *Biomolecules* **11**, 1671 (2021).
50. Alexander, N. R. et al. Extracellular matrix rigidity promotes invadopodia activity. *Curr. Biol.* **18**, 1295–1299 (2008).
51. White, E. S. Lung extracellular matrix and fibroblast function. *Ann. Am. Thorac. Soc.* **12**, S30–S33 (2015).
52. Burgstaller, G. et al. The instructive extracellular matrix of the lung: basic composition and alterations in chronic lung disease. *Eur. Respir. J.* **50**, 1601805 (2017).
53. Zhou, Y. et al. Inhibition of mechanosensitive signaling in myofibroblasts ameliorates experimental pulmonary fibrosis. *J. Clin. Invest.* **123**, 1096–1108 (2013).
54. Zhao, P. et al. The CD44s splice isoform is a central mediator for invadopodia activity. *J. Cell. Sci.* **129**, 1355–1365 (2016).
55. Geng, Y. et al. PD-L1 on invasive fibroblasts drives fibrosis in a humanized model of idiopathic pulmonary fibrosis. *JCI Insight* **4**, e125326 (2019).
56. Ahluwalia, N. et al. Fibrogenic lung injury induces non-cell-autonomous fibroblast invasion. *Am. J. Respir. Cell. Mol. Biol.* **54**, 831–842 (2016).
57. Jiang, D. et al. Inhibition of pulmonary fibrosis in mice by CXCL10 requires glycosaminoglycan binding and syndecan-4. *J. Clin. Invest.* **120**, 2049–2057 (2010).
58. Liu, X. et al. HER2 drives lung fibrosis by activating a metastatic cancer signature in invasive lung fibroblasts. *J. Exp. Med.* **219**, e20220126 (2022).
59. Spinardi, L. et al. A dynamic podosome-like structure of epithelial cells. *Exp. Cell Res.* **295**, 360–374 (2004).
60. Michopoulou, A. et al. A novel mechanism in wound healing: Laminin 332 drives MMP9/14 activity by recruiting syndecan-1 and CD44. *Matrix Biol.* **94**, 1–17 (2020).
61. Jaeger, B. et al. Airway basal cells show a dedifferentiated KRT17(high) phenotype and promote fibrosis in idiopathic pulmonary fibrosis. *Nat. Commun.* **13**, 5637 (2022).
62. Selman, M. et al. TIMP-1, -2, -3, and -4 in idiopathic pulmonary fibrosis. A prevailing nondegradative lung microenvironment? *Am. J. Physiol. Lung Cell. Mol. Physiol.* **279**, L562–L574 (2000).
63. Espindola, M. S. et al. Differential responses to targeting matrix metalloproteinase 9 in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **203**, 458–470 (2021).
64. Yu, Q. & Stamenkovic, I. Cell surface-localized matrix metalloproteinase-9 proteolytically activates TGF- β and promotes tumor invasion and angiogenesis. *Genes Dev.* **14**, 163–176 (2000).
65. Ramírez, G. et al. Absence of Thy-1 results in TGF- β induced MMP-9 expression and confers a profibrotic phenotype to human lung fibroblasts. *Lab. Invest.* **91**, 1206–1218 (2011).
66. Bormann, T. et al. Role of matrix metalloproteinase-2 and MMP-9 in experimental lung fibrosis in mice. *Respir. Res.* **23**, 180 (2022).
67. Ahangari, F. et al. Saracatinib, a selective src kinase inhibitor, blocks fibrotic responses in preclinical models of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **206**, 1463–1479 (2022).
68. Tzouvelekkis, A. et al. SH2 domain-containing phosphatase-2 is a novel antifibrotic regulator in pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **195**, 500–514 (2017).
69. Papadopoulou, A. & Kleatsas, D. Human lung fibroblasts prematurely senescent after exposure to ionizing radiation enhance the growth of malignant lung epithelial cells in vitro and in vivo. *Int. J. Oncol.* **39**, 989–999 (2011).
70. Díaz, B. Invadopodia detection and gelatin degradation assay. *Bio. Protoc.* **3**, e997 (2013).
71. DeQuach, J. A. et al. Simple and high yielding method for preparing tissue specific extracellular matrix coatings for cell culture. *PLoS ONE* **5**, e13039 (2010).
72. Kawai, N. et al. Induction of lung-like cells from mouse embryonic stem cells by decellularized lung matrix. *Biochem. Biophys. Rep.* **15**, 33–38 (2018).
73. Rothberg, J. M. et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**, 348–352 (2011).
74. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e1821 (2019).
75. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e3529 (2021).
76. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
77. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
78. Fanidis, D. & Moulos, P. Integrative, normalization-insusceptible statistical analysis of RNA-Seq data, with improved differential expression and unbiased downstream functional analysis. *Brief. Bioinform.* **22**, bbaa156 (2020).
79. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
80. Risso, D., Schwartz, K., Sherlock, G. & Dudoit, S. GC-content normalization for RNA-seq data. *BMC Bioinform.* **12**, 480–480 (2011).
81. The Gene Ontology Consortium. The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Res.* **49**, D325–D334 (2021).
82. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: J. Integr. Biol.* **16**, 284–287 (2012).
83. Rani, J., Shah, A. R. & Ramachandran, S. pubmed.mineR: an R package with text-mining algorithms to analyse PubMed abstracts. *J. Biosci.* **40**, 671–682 (2015).
84. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
85. Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* **29**, 1363–1375 (2019).
86. Barbayianni, I. et al. Dataset. <https://doi.org/10.5281/zenodo.8296510> (2023).

Acknowledgements

We would like to thank Vassia Papadaki (Kafasla lab, BSRC Fleming) for support on confocal imaging, and Dimitris Kleatsas (NCSR Demokritos) for NHLF clones and advice on optimal NHLF culture conditions.

This research was co-financed by Greece and the European Union (European Social Fund—ESF) through the Operational Program «Human Resources Development, Education and Lifelong Learning» in the context of the project “Strengthening Human Resources Research Potential via Doctorate Research” (MIS-5000432), implemented by the State Scholarships Foundation (Fellowship to IB). Research was further partly supported through the Hellenic Foundation for Research and Innovation (HFRI) under the “2nd Call for HFRI Research Projects to support Faculty Members & Researchers” (#3565 to VA). Human studies were supported by the NIH NHLBI grants R01HL127349, R01HL141852, U01HL145567, and UH2HL123886 to NK. The funders had no role in the design of the study, or in the collection, analyses, or

interpretation of data, or in the writing of the manuscript, or in the decision to publish the results.

Author contributions

I.B. and P.K. performed most presented experiments and analyzed the relative data, assisted by E.-D.N., D.N., A.G., and E.T.; A.T. performed human studies supervised by R.H., N.K. and V.A.; D.F. performed in silico data re-analysis and supervised all statistical analyses. V.H. and P.H. performed RNAseq. K.A., B.C. and N.K. provided resources. V.A. conceived and coordinated the project. I.B., P.K., D.F. and V.A. wrote the paper, which was edited by V.A. and critically commented by all co-authors.

Competing interests

A.T. has received fees for speaking and/or organizing education from AstraZeneca, Menarini, Boehringer Ingelheim, Chiesi, Hoffmann-La Roche, Ltd., GlaxoSmithKline and Elpen, for consulting from Boehringer Ingelheim, Pfizer, Gilead, Hoffmann-La Roche, Ltd., GlaxoSmithKline, and has received research funding, including institutional funding, from Boehringer Ingelheim, Chiesi, Hoffmann-La Roche, Ltd., GlaxoSmithKline and Astra Zeneca, outside the submitted work. B.C. has received fees for speaking and/or organizing education from Apellis, Astra Zeneca, BMS, Boehringer Ingelheim, Novartis, Roche and Sanofi, for consulting fees from Apellis, BMS, Boehringer Ingelheim and Sanofi, and has received research funding from Boehringer Ingelheim, outside the submitted work. N.K. is a scientific founder at Thyron, served as a consultant to Biogen Idec, Boehringer Ingelheim, Third Rock, Pliant, Samumed, NuMedii, Theravance, LifeMax, Three Lake Partners, Optikira, Astra Zeneca, RohBar, Veracyte, Augmanity, CSL Behring, Galapagos and Thyron over the last 3 years, reports Equity in Pliant and Thyron, and grants from Veracyte, Boehringer Ingelheim, BMS and non-financial support from MiRagen and Astra Zeneca, outside the submitted work. Other authors declare that they have no conflict of interest.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-41614-x>.

Correspondence and requests for materials should be addressed to Vassilis Aidinis.

Peer review information *Nature Communications* thanks John Lewis, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

MAP3K8 Regulates Cox-2-Mediated Prostaglandin E₂ Production in the Lung and Suppresses Pulmonary Inflammation and Fibrosis

Η MAP3K8 ρυθμίζει την Cox-2 διαμεσολαβούμενη παραγωγή προσταγλανδίνης E₂ στον πνεύμονα και καταστέλλει την πνευμονική φλεγμονή και ίωση

Η ιδιοπαθής πνευμονική ίωση (IPF) χαρακτηρίζεται από άφθονη εναπόθεση συστατικών της εξωκυττάριας μήτρας οδηγώντας σε επιδείνωση της πνευμονικής αρχιτεκτονικής και των λειτουργιών της αναπνοής. Οι προϋνωτικοί μηχανισμοί ελέγχονται από πολλαπλά ρυθμιστικά μόρια, συμπεριλαμβανομένων των MAPKs, τα οποία με τη σειρά τους ρυθμίζονται από πολλαπλούς καταρράκτες φωσφορυλίωσης. Η MAP3K8 είναι μια MAPK κινάση κινάση που έχει προταθεί πως ρυθμίζει πλειοτροπικώς πολλαπλά παθολογικά μονοπάτια στο πλαίσιο της ίωσης και του καρκίνου· εντούτοις πιθανός ρόλος της στην παθογένεση της IPF δεν έχει ερευνηθεί. Σε αυτή την αναφορά τα επίπεδα μεταφραστικού RNA του *MAP3K8* βρέθηκαν μειωμένα στους πνεύμονες ασθενών της IPF και σε αυτούς ποντικών, μετά την επαγωγή πνευμονικής ίωσης χρήση μπλεομυκίνης (BLM). Ευρεία γενετική διαγραφή του *Map3k8* σε ποντικούς επιδείνωσε την μοντελοποιημένη ασθένεια, και πειράματα μεταφοράς μυελού των οστών έδειξαν πως ενώ οι ρυθμιστικοί μηχανισμοί της MAP3K8 είναι ενεργοί τόσο σε αιμοποιητικά όσο και μη αιμοποιητικά κύτταρα, ο ρόλος τους στα πρώτα είναι περισσότερο κυρίαρχος. Ειδική διαγραφή του *Map3k8* από μακροφάγα βρέθηκε να είναι επαρκής για την επιδείνωση της ασθένειας, επιβεβαιώνοντας με αυτόν τον τρόπο τον δεσπόζοντα ρόλο τους στις αποκρίσεις της πνευμονικής ίωσης και προτείνοντας τον κύριο ρόλο της Map3k8 στην ομοιοστάση των εκτελεστικών πνευμονικών λειτουργιών τους. Επίσης, ανεπάρκεια του *Map3k8* βρέθηκε να σχετίζεται με μειωμένη έκφραση του Cox-2, και επακόλουθη μειωμένη παραγωγή PGE₂ στους πνεύμονες· αντιστοίχως εξωγενής χορήγηση PGE₂ μείωσε την φλεγμονή και ανέστρεψε το επιδεινωμένο ινωτικό προφίλ των *Map3k8*^{-/-} ποντικών. Ως εκ τούτου, η MAP3K8 έχει κεντρικό ρόλο στην ρύθμιση φλεγμονωδών αποκρίσεων, στην Cox-2-διαμεσολαβούμενη παραγωγή PGE₂ στον πνεύμονα, ενώ ύφεση της έκφρασης της είναι καίριας σημασίας για την ανάπτυξη πνευμονικής ίωσης.

Luminex
complexity simplified.



Reimagine your discoveries

Amnis[®] ImageStream[™] Mk II and
FlowSight[™] Imaging Flow Cytometers

Learn more >



MAP3K8 Regulates Cox-2–Mediated Prostaglandin E₂ Production in the Lung and Suppresses Pulmonary Inflammation and Fibrosis

This information is current as of September 26, 2022.

Markella Zannikou, Ilianna Barbayianni, Dionysios Fanidis, Theodora Grigorakaki, Evlalia Vlachopoulou, Dimitris Konstantopoulos, Maria Fousteri, Ioanna Nikitopoulou, Anastasia Kotanidou, Eleanna Kaffe and Vassilis Aidinis

J Immunol 2021; 206:607-620; Prepublished online 18 December 2020;

doi: 10.4049/jimmunol.2000862

<http://www.jimmunol.org/content/206/3/607>

Supplementary Material <http://www.jimmunol.org/content/suppl/2020/12/17/jimmunol.2000862.DCSupplemental>

References This article **cites 66 articles**, 19 of which you can access for free at: <http://www.jimmunol.org/content/206/3/607.full#ref-list-1>

Why *The JI*? [Submit online.](#)

- **Rapid Reviews! 30 days*** from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Fast Publication!** 4 weeks from acceptance to publication

**average*

Subscription Information about subscribing to *The Journal of Immunology* is online at: <http://jimmunol.org/subscription>

Permissions Submit copyright permission requests at: <http://www.aai.org/About/Publications/JI/copyright.html>

Email Alerts Receive free email-alerts when new articles cite this article. Sign up at: <http://jimmunol.org/alerts>

The Journal of Immunology is published twice each month by
The American Association of Immunologists, Inc.,
1451 Rockville Pike, Suite 650, Rockville, MD 20852
Copyright © 2021 by The American Association of
Immunologists, Inc. All rights reserved.
Print ISSN: 0022-1767 Online ISSN: 1550-6606.



MAP3K8 Regulates Cox-2–Mediated Prostaglandin E₂ Production in the Lung and Suppresses Pulmonary Inflammation and Fibrosis

Markella Zannikou,* Ilianna Barbayianni,* Dionysios Fanidis,* Theodora Grigorakaki,* Evlalia Vlachopoulou,* Dimitris Konstantopoulos,[†] Maria Fousteri,[†] Ioanna Nikitopoulou,[‡] Anastasia Kotanidou,^{‡,§} Eleanna Kaffe,* and Vassilis Aidinis*

Idiopathic pulmonary fibrosis (IPF) is characterized by exuberant deposition of extracellular matrix components, leading to the deterioration of lung architecture and respiratory functions. Profibrotic mechanisms are controlled by multiple regulatory molecules, including MAPKs, in turn regulated by multiple phosphorylation cascades. MAP3K8 is an MAPK kinase kinase suggested to pleiotropically regulate multiple pathogenic pathways in the context of inflammation and cancer; however, a possible role in the pathogenesis of IPF has not been investigated. In this report, *MAP3K8* mRNA levels were found decreased in the lungs of IPF patients and of mice upon bleomycin-induced pulmonary fibrosis. Ubiquitous genetic deletion of *Map3k8* in mice exacerbated the modeled disease, whereas bone marrow transfer experiments indicated that although MAP3K8 regulatory functions are active in both hematopoietic and nonhematopoietic cells, *Map3k8* in hematopoietic cells has a more dominant role. Macrophage-specific deletion of *Map3k8* was further found to be sufficient for disease exacerbation thus confirming a major role for macrophages in pulmonary fibrotic responses and suggesting a main role for *Map3k8* in the homeostasis of their effector functions in the lung. *Map3k8* deficiency was further shown to be associated with decreased *Cox-2* expression, followed by a decrease in PGE₂ production in the lung; accordingly, exogenous administration of PGE₂ reduced inflammation and reversed the exacerbated fibrotic profile of *Map3k8*^{-/-} mice. Therefore, MAP3K8 has a central role in the regulation of inflammatory responses and Cox-2–mediated PGE₂ production in the lung, and the attenuation of its expression is integral to pulmonary fibrosis development. *The Journal of Immunology*, 2021, 206: 607–620.

Idiopathic pulmonary fibrosis (IPF) is an interstitial lung disease with a dismal prognosis, characterized by fibroblast foci and exuberant deposition of extracellular matrix components, leading to the distortion of lung architecture and the deterioration of respiratory functions. The prevailing working

hypothesis suggests that the mechanisms driving IPF reflect abnormal, deregulated wound healing in response to persistent, environmentally imposed epithelial damage in genetically predisposed individuals of an older age (1, 2). Although the only effective treatment remains lung transplantation, pirfenidone (3) and nintedanib (4) were found to delay disease's progression and constitute the current standard of care. Nintedanib is a small molecule that inhibits receptor tyrosine-kinases (RTKs; such as FGFR, VEGFR, and PDGFR), membrane receptors that activate multiple, and frequently overlapping, cellular signaling pathways, including MAPK pathways (4, 5). Although much less is known on the mode of action of pirfenidone, it is generally considered as an antifibrotic and anti-inflammatory agent, regulating the expression and activity of TGF- β and TNF- α that both, variably, rely on MAPK signaling to exert their effector functions (3, 5).

MAPKs, such as ERKs, JNKs, and p38 MAPKs, are protein Ser/Thr kinases that phosphorylate many intracellular targets including cytoskeletal elements, membrane transporters, as well as other kinases and transcription factors, resulting in the regulation of a wide array of cellular functions, including stress response, cell growth, differentiation, proliferation, and apoptosis (6). Aberrant MAPK signaling has been linked to the pathogenesis of many inflammatory, metabolic, or malignant diseases, highlighting the importance of dissecting the regulatory mechanisms governing their activation (6). MAPKs are activated from MAPK kinases (MAP2Ks), in turn activated by MAP2K kinases (MAP3Ks), which provide stimulus- and cell-specific signaling contexts for cellular responses to extracellular stimuli such as peptide growth factors, cytokines, hormones, as well as to endoplasmic reticulum and oxidative stress (6), all inherently linked to IPF pathogenesis.

*Institute of Bio-Innovation, Biomedical Sciences Research Center Alexander Fleming, 16672 Athens, Greece; [†]Institute for Fundamental Biomedical Research, Biomedical Sciences Research Center Alexander Fleming, 16672 Athens, Greece; [‡]GP Livanos and M Simou Laboratories, National and Kapodistrian University of Athens, Evangelismos Hospital, 106 76 Athens, Greece; and [§]First Department of Critical Care and Pulmonary Services, Medical School, National and Kapodistrian University of Athens, Evangelismos Hospital, Athens 106 76, Greece

ORCID: 0000-0003-1452-2056 (M.Z.); 0000-0002-4053-2090 (D.F.); 0000-0003-1729-5899 (M.F.); 0000-0001-9531-7729 (V.A.).

Received for publication July 28, 2020. Accepted for publication November 17, 2020.

This work was supported by an “excellence” grant from the General Secretariat of Research and Technology, Hellenic Ministry of Education, lifelong learning and religious affairs (Aristia II; 3311).

Address correspondence and reprint requests to Dr. Vassilis Aidinis, Biomedical Sciences Research Center Alexander Fleming, 34 Fleming Street, 16672 Athens, Greece. E-mail address: v.aidinis@fleming.gr

The online version of this article contains supplemental material.

Abbreviations used in this article: AA, arachidonic acid; BALF, bronchoalveolar lavage fluid; BLM, bleomycin; B2M, β 2-microglobulin; Cox-2, cyclooxygenase-2; dmPGE₂, 16,16-dimethyl PGE₂; GEO, Gene Expression Omnibus; HC, hematopoietic cell; IM, interstitial macrophage; IPF, idiopathic pulmonary fibrosis; MAP2K, MAPK kinase; MAP3K, MAP2K kinase; MAP3K8, MAPK kinase kinase 8; Mo-AM, monocyte-derived alveolar macrophage; nHC, nonhematopoietic cell; PCA, principal component analysis; Q-RT-PCR, quantitative RT-PCR; rma, robust multi-array average; scRNAseq, single-cell RNA-sequencing; SSC, side scatter; SSC-A, SSC area; TPL2, tumor progression locus 2; wt, wild-type.

Copyright © 2021 by The American Association of Immunologists, Inc. 0022-1767/21/\$37.50

MAPK kinase kinase 8 (MAP3K8; tumor progression locus 2 [TPL2]; cancer Osaka thyroid [COT] oncogene) is an Ser/Thr MAP3K, at the crossroad of many different signaling pathways. TPL2 was first identified by virtue of its oncogenic transforming activity in cells, and is now incriminated for the regulation of multiple pathogenic pathways in the context of inflammation and cancer (7). MAP3K8 has been reported to promote pleiotropic and even opposing effects, depending on the cell type and the tissue microenvironment. Although MAP3K8 is generally considered as a proinflammatory molecule and an oncogene, many reports have indicated an opposite role in limiting inflammatory and fibrotic responses and suppressing carcinogenesis (7). In the lung, in which *Tpl2/Map3k8* is highly expressed, *Map3k8* ubiquitous deficiency was shown to exacerbate eosinophilic inflammation when challenged with OVA (8) and to promote urethane-induced lung carcinogenesis (9). Because pulmonary fibrosis is a major risk factor for the development of lung cancer (10), and because different Map3k8-regulated functions are relative with many cellular mechanisms governing the pathogenesis of pulmonary fibrosis, in this report, we examined a possible role of Map3k8 in regulating pathogenetic responses in pulmonary fibrosis.

Materials and Methods

Mice

All mice were bred in the C57Bl6/J background for over 10 generations under specific pathogen-free conditions at the local animal facilities. Mice were housed at 55 ± 5% humidity, 20–22°C under a 12-h light-dark cycle; food and water were given ad libitum. All experimentation, conforming to the Animal Research: Reporting In Vivo Experiments (ARRIVE) guidelines, was approved by the internal Institutional Review Board (no. 95), as well as by the Veterinary Service and Fishery Department of the local governmental prefecture (no. 1121). The generation and genotyping instructions of *LysM-Cre* (11), *Map3k8^{fl/fl}*, and *Map3k8^{-/-}* mice (12) have been described previously. All randomly assigned experimental groups consisted of littermate mice. The health status of the mice was monitored once per day; no unexpected deaths were observed; all measures were taken to minimize animal suffering and distress.

Bleomycin-induced pulmonary fibrosis

Pulmonary fibrosis was induced by a single intratracheal instillation (with a MicroSprayer aerosolizer) of bleomycin (BLM) (3.2 U/Kg; Nippon Kayaku) into anesthetized (i.p.; xylazine, ketamine, and atropine, 10, 100, and 0.05 mg/kg, respectively) 6- to 8-wk-old male and female mice as previously reported (13–15). It should be noted that beyond dosing and route of administration, the timing, severity, and resolution of BLM-induced fibrosis depends on a number of parameters, such as the specific genetic background (i.e., C57bl6 J or N, genetic drift, vendor) and local health status. The dose used in this study was selected upon prior extensive local testing to minimize lethality (for ethical and practical reasons) while inducing a solid fibrotic profile, analyzed with a plethora of readout assays as previously reported (13) and as described below.

Bone marrow transplantation

Bone marrow transplantation was performed as previously described (16); *Map3k8^{-/-}* and wild-type (wt) mice were irradiated once with 1000 rad (J. L. Shepherd Irradiator). Irradiated mice were injected i.v. through the tail vein with 5×10^6 bone marrow cells from male wt or *Map3k8^{-/-}* mice. All the irradiated nontransplanted control mice died within 14 d from the time of irradiation, whereas all implanted mice survived, confirming efficient transplantation and reconstitution of the hemopoietic system.

PGE₂ administration

The 16,16-dimethyl PGE₂ (dmPGE₂; Cayman Chemical, Ann Arbor, MI) was dissolved in methyl acetate, which was then evaporated under a nitrogen stream; dmPGE₂ was immediately dissolved in nitrogen-purged ethanol and kept as a stock solution at a concentration of 0.5 mg/ml. Immediately before in vivo administration, the stock dmPGE₂ solution was diluted with saline and kept on ice. dmPGE₂ was administered starting 1 d before BLM via two daily i.p. injections at a concentration of 10 µg/kg.

Measurement of respiratory functions

Respiratory mechanics, recently shown to be an accurate surrogate marker of disease development (13), were analyzed in tracheostomized mice with the FlexiVent ventilator system (SCIREQ) following manufacturer instructions, as previously reported (13).

Bronchoalveolar lavage fluid isolation and measurements

Bronchoalveolar lavage fluid (BALFs) were obtained by lavaging the lungs of tracheostomized mice with 1 ml of 0.9% sterile sodium chloride (three times). After centrifugation at $100 \times g$ for 10 min (4°C), the first BALF supernatant was stored at -80°C for protein and collagen content determination, and cell pellets from total BALF volume were counted with a hemacytometer after being stained with 0.4% trypan blue solution or used for FACS analysis.

Total protein levels were assessed with the Bradford assay according to manufacturer's instructions (Bio-Rad Laboratories, Hercules, CA). Absorbance values were converted in mg/ml using a BSA standard curve (BSA 0–2 mg/ml). Total soluble collagen was quantified using the Sirius Red assay protocol; briefly, 50 µl of BALF samples, diluted in 0.5 M acetic acid, were incubated for 30 min with Sirius Red at room temperature (direct red 80; 120 µg/ml in 0.5 M acetic acid). After centrifugation ($12,000 \times g$ for 10 min), the absorbance of supernatant was read at 540 nm, and values were converted in micrograms per milliliter according to a standard curve with collagen type I from rat tail (0–500 µg/ml).

For FACS analysis, BALF cells were centrifuged at 1200 rpm at 4°C and incubated with Fc Receptor Binding Inhibitor C16/32 (eBioscience) for 15 min on ice. Cells were then incubated for 30 min in FACS buffer containing manufacturers' suggested dilutions of fluorescently labeled mAbs. The following Abs were used for analysis: anti-CD11b-PE, anti-F4/80-PE, anti-CD4-AF700, anti-CD8-allophycocyanin, anti-B220-PerCP-Cy5.5, and anti-Gr1-FITC (all BioLegend, San Diego, CA). Live hematopoietic cells (HCs) were first gated empirically by forward scatter versus side scatter characteristics (SSC). Cells in the hematopoietic gate were then interrogated for surface immunophenotypic markers such as CD8, CD4 (T cells), or B220 (B cells). Neutrophils were recognized as non-autofluorescent highly granular (SSC^{hi}) cells and, within this gate, were defined as cells F4/80⁺-Gr1^{high} (R1) or by gating for CD11b⁺ (CD11b versus SSC area [SSC-A]) cells and these were further gated for Gr1^{high} (neutrophils Gr1 versus SSC-A). CD11b/Gr1^{high} cells had multilobed nuclei typical of granulocytes, whereas CD11b/Gr1^{mid} and CD11b/Gr1^{low} cells had ovoid nuclei typical of monocytes/macrophages. Macrophages were identified as large autofluorescent cells F4/80⁺Gr1^{mid/low} (R2) or by gating for CD11b⁺ (CD11b versus SSC-A) cells and these were further gated for Gr1^{low} (macrophages-monocytes) (Gr1 versus SSC-A). All acquisition was performed using a BD FACSymphony flow cytometer. Data analysis was performed with the FlowJo software.

Lung histology

The right lung tissues were fixed in 10% v/v neutral buffered formalin and embedded in paraffin. Four-micrometer lung sections were prepared and stained with H&E with standard protocols. For Sirius Red-Fast Green staining, tissue sections were deparaffinized and stained successively with Fast Green 0.04% and Sirius Red 0.1%/Fast Green 0.04% dissolved in picric acid. Finally, sections were mounted with distyrene plasticizer xylene. Lung tissue imaging was performed using a Nikon Eclipse E800 microscope (Nikon, Shinagawa-ku, Japan) attached to a Q Imaging EXI Aqua Digital Camera, using the Q-Capture Pro 7 software.

Quantitative RT-PCR

Total RNA was extracted from the left lung lobe from each mouse using the Tri Reagent (Molecular Research Center), followed by DNase treatment (RQ1 RNase-free DNase; Promega, Madison, WI) according to the respective manufacturers' instructions. cDNA was synthesized from 2 µg of total RNA in a 20-µl reaction using Moloney murine leukemia virus reverse transcriptase (Promega). Quantitative RT-PCR (Q-RT-PCR) was performed using SoFast EvaGreen Supermix in a Bio-Rad CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories). Values were normalized to β2-microglobulin (B2M). Primers used, as well as the product size (bp) were as follows: *coll1a1* (forward: 5'-CTA CTA CCG GGC CGA TGA TG-3'; reverse: 5'-CGA TCC AGT ACT CTC CGC TC-3'; 188 bp), *map3k8* (forward: 5'-TTA GCC CAA GAC ATG AAG AC-3'; reverse: 5'-ACT CAG CAA TGT TCT CAT GC-3'; 117 bp), and *b2m* (forward: 5'-TTC TGG TGC TTG TCT CAC TGA-3'; reverse: 5'-CAG TAT GTT CGG CTT CCC ATTC-3'; 104 bp). The annealing temperature for all primers was 58°C. Values were calculated according to the $2^{-\Delta\Delta\text{cycle threshold}}$ method.

Mass spectrometry

As the estimated half-life of PGE₂ in vivo is less than 15 s (17), we quantified PGE₂ through its metabolites, as recently reported (18). Briefly, 300 μ l of BALFs were deproteinized with acetone. After vigorous mixing for 4 min and centrifugation at 2000 \times g for 10 min at 4°C, samples were transferred to a clean 15-ml glass vial, mixed with 800 μ l of hexane by vigorous mixing for 30 s and centrifuged for 10 min at 2000 \times g at 4°C. The lower phase was acidified to pH 3.5 with formic acid and then mixed with chloroform. After three vigorous mixing for 30 s and centrifugation for 10 min at 2000 \times g at 4°C, the lower chloroform phase was kept for 15 min at -80°C to separate any residual from the upper phase. Samples extracted in chloroform were evaporated to dryness using a speed-vac concentrator (model no. SC110-120; Savant Systems, San Diego, CA) and redissolved in 100 μ l of methanol/100 mM ammonium acetate pH 8.5 (9:1). The dmPGE₂ (Cayman Chemicals) was used as an internal control in each sample from the beginning of the above extraction procedure at a final concentration of 2.5 ng/ml. Samples were analyzed by direct infusion in an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Waltham, MA) with electrospray ionization in negative mode. Capillary temperature was 275°C, spray voltage was 3.5 kV, sheath gas was set at 40 U, and sweep gas was at 8 U. The resolution was 100 K, providing high accuracy for prostanoid measurements. Precursor ion masses were used for prostanoid profiling, whereas lipid identity was confirmed by precursor ion fragmentation using collision induced dissociation. Precursor m/z 297.1530 was used for 13,14-dihydro-15-keto-tetranor-PGE₂ and tetranorPGE₁, stable metabolites of PGE₂. Precursor m/z Precursor m/z 325.2021 was used for metabolites 2,3-dinor-11b-PGF2 α , 2,3-dinor-PGE₁, dinor-PGF2 α , and 2,3-dimensionalinor-8-iso-PGF2 α .

Microarray and scRNAseq data reanalysis

The raw microarray datasets GSE32537 and GSE47460 (that contains data from two different platforms, GPL6480 and GPL14550, which were processed separately) were downloaded from Gene Expression Omnibus (GEO), and intensity values were background corrected with subtraction or normexp methods implemented into limma (19) and oligonucleotide (20) Bioconductor packages, respectively, and robust multi-array average (rma) normalized. Afterwards, sample outliers were removed based on a principal component analysis (PCA) plot created with arrayQualityMetrics Bioconductor package (21). Control probes, as well as probes mapping to more than one HUGO gene symbol were removed from further analysis. Intensity values were then summarized at the gene level using a weighted average value of all probes/transcript clusters representing each gene (weights sum up to the unit). Subsequent differential expression analysis was performed using the moderated *t* test statistics algorithm provided by limma R package.

scRNAseq GSE122960 dataset was processed with Seurat R package version 3.1.2 (22). Adaptation of the original reported code provided in this study was used to reanalyze the GSE122960 hierarchical feature-barcode matrices. Because of the Seurat update since the original GSE122960 publication (23), samples integration was performed with a two-step application of the standard Seurat v3 integration pipeline: integration of the donor and IPF samples separately prior to donor-IPF samples final integration. All analyses were performed on R version 3.6.2, and the datasets were retrieved from the GEO repository.

Statistical analysis

Statistical significance was assessed with the Prism (GraphPad) software, as detailed at each figure legend.

Results

Decreased pulmonary MAP3K8 mRNA expression in pulmonary fibrosis

To examine a possible role of Map3k8 in the regulation of pathogenic signal transduction pathways during fibrogenesis, we first sought to quantify its mRNA expression levels during the development of BLM-induced pulmonary inflammation and fibrosis. To this end, BLM (3.2 U/Kg) was intratracheally administered to 6- to 8-wk-old C57Bl6/J mice, which were then sacrificed 7, 14, and 21 d post-BLM administration, roughly corresponding to the (postacute) inflammatory, fibrotic, and resolution phases of disease development. As expected, surviving (Fig. 1A) mice that received BLM lost weight (Fig. 1B), accompanied by a gradual increase in pulmonary edema and inflammation, as indicated from the total

protein concentration and the infiltrating cell numbers of the BALF (Fig. 1C, 1D, respectively). *Coll1a1* mRNA expression, as determined with Q-RT-PCR in lung tissue from the same mice, was also found gradually increasing post-BLM (Fig. 1E), as also reflected at the soluble collagen levels in the BALF (Fig. 1F), determined with the Sirius Red assay. Accordingly, histopathological analysis of the lungs of mice post-BLM administration indicated the increasing presence of peribronchiolar and parenchymal fibrotic regions (Fig. 1G), resulting in impaired respiratory mechanics, such as resistance, tissue elasticity, and static compliance, as measured with FlexiVent (Fig. 1H–J). All disease signs peaked 14 d post-BLM, subsiding at 21 d (Fig. 1C–G). Q-RT-PCR mRNA analysis (in the same samples as in Fig. 1E) indicated a gradual decrease in pulmonary *Map3k8* expression in lung tissue in inverse correlation with disease development (Fig. 1K).

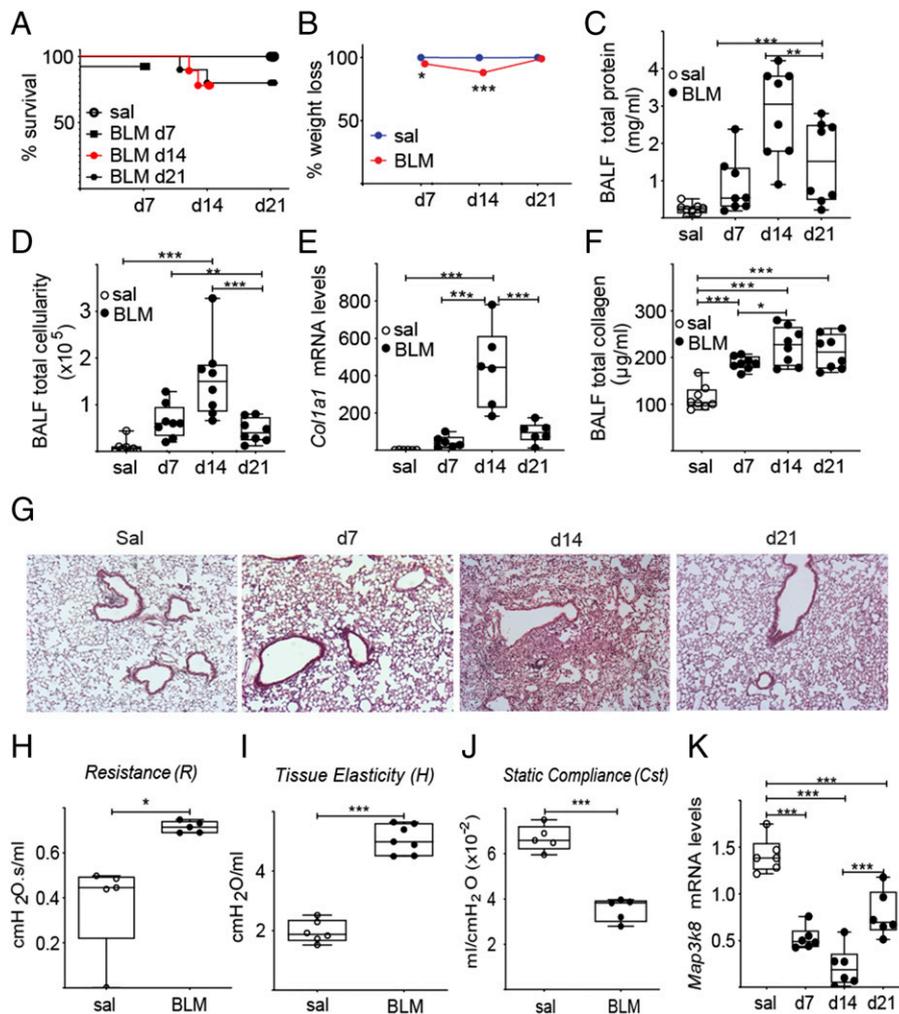
To translate the findings into the human disease, we performed in silico reanalysis of the three largest publicly available expression profiling datasets (GSE32537 and GSE47460) at GEO interrogating differential expression in the lung tissue of 119, 122, and 38 IPF patients in comparison with 50, 91, and 17 controls, respectively (24, 25). Raw data were background corrected and rma normalized, and outliers were removed based on PCA plots (Fig. 2A, 2B, 2D–G). Differential expression analysis indicated a modest but highly statistical significant decrease of *MAP3K8* mRNA levels in IPF (\log_2 FC/FDR-corrected *p* values for the different datasets GSE32537, GSE47460-GPL6480, and GSE47460-GPL14550 are respectively $-0.459/4.65 \times 10^{-10}$, $-0.658/1.06 \times 10^{-4}$, and $-0.668/1.46 \times 10^{-13}$; Fig. 2), in agreement with the modeled disease (Fig. 1). Thus, the development of pulmonary fibrosis, in both humans and mice, is accompanied with a transcriptional (or posttranscriptional) downregulation of *MAP3K8* mRNA levels in the lung.

As *MAP3K8* is widely expressed in different pulmonary cell types exerting pleiotropic effects (7), we next visualized *MAP3K8* mRNA expression in single pulmonary cells in the normal or fibrotic human lung, via the web-tool available at nupulmonary.org/resources and reanalysis of the relative single-cell RNA-sequencing (scRNAseq) dataset GSE122960 (23). As evident in Supplemental Fig. 1A, *MAP3K8* is expressed at both nonhematopoietic (nHCs) and HCs, and most notably in monocytes/macrophages. More specifically, reduced *MAP3K8* expression was detected in the *SPPI1*⁺ macrophage subcluster 1, which most likely correspond to monocyte-derived macrophages ($p = 6.11 \times 10^{-24}$; Supplemental Fig. 1B). In addition, the same data also suggest *MAP3K8* downregulation in the (aSMA⁺) myofibroblast cell cluster ($p = 0.0002$; Supplemental Fig. 1C), further corroborating a transcriptional downregulation of *MAP3K8* expression in IPF.

Map3k8, especially in macrophages, has a protective role in BLM-induced pulmonary fibrosis

To validate and dissect a possible Map3k8 role in pulmonary fibrosis, BLM was administered to *Map3k8* ubiquitous knockout mice (*Map3k8*^{-/-}) (12) and wt littermate controls. *Map3k8*^{-/-} mice present with no apparent pulmonary phenotype or impaired respiratory functions upon healthy conditions (Fig. 3; Sal control groups). BLM administration to *Map3k8*^{-/-} mice resulted in increased lethality (Fig. 3A) and in greater weight loss in the surviving mice (Fig. 3B), suggesting increased systemic disease burden in comparison with littermate wt mice. Pulmonary edema and inflammation were also found significantly increased in surviving *Map3k8*^{-/-} mice (Fig. 3C, 3D, respectively), accompanied by increased collagen expression (Fig. 3E, 3F). Accordingly, tissue fibrosis was notably expanded (Fig. 3G), resulting in further

FIGURE 1. Decreased pulmonary *Map3k8* expression upon BLM-induced pulmonary fibrosis. **(A)** Kaplan Meyer survival and **(B)** weight loss of wt C57Bl6/J mice post-BLM administration. **(C)** Total protein concentration in BALF, as determined with the Bradford assay. **(D)** Inflammatory cell numbers in BALFs, as counted with a hemacytometer. **(E)** *Col1a1* mRNA levels in whole-lung tissue were determined with Q-RT-PCR analysis; values were normalized to the expression of B2M. **(F)** Soluble collagen in BALFs was determined with the Sirius Red assay. **(G)** Representative H&E-stained sections of murine lungs at 7, 14, and 21 d post-BLM (original magnification $\times 10$). **(H–J)** Indicated respiratory functions were measured with FlexiVent. **(K)** *Map3k8* mRNA levels were quantified in the same samples as in (E) with Q-RT-PCR. Statistical significance was assessed with two-way ANOVA with Bonferroni post hoc correction (A–F) and unpaired *t* test with Welch correction (two-tailed; H–J). Data are presented as means \pm SEM ($n = 5–7$; cumulative results from two separate experiments). * $p = 0.05$, ** $p = 0.01$, *** $p = 0.001$.



deterioration of respiratory functions upon BLM administration (Fig. 3H–J).

As *Map3k8* is expressed from both nHCs and HCs (Supplemental Fig. 1), we next interrogated the relative contribution of *Map3k8* expression from nHCs and HCs in disease protection. To that end, wt and *Map3k8*^{-/-} mice were irradiated to abolish HCs, followed by injections of bone marrow cells isolated from wt or *Map3k8*^{-/-} mice to reconstitute the immune system (Fig. 4A). The generated chimeric mice bearing HCs with the genetic background of the donor (Fig. 4B), were then administered BLM and disease severity was assessed 14 d post-BLM, at the peak of the disease. *Map3k8* deficiency in HCs resulted in a bigger effect in the systemic manifestations of the disease, as reflected in lethality (Fig. 4C) and weight loss (Fig. 4D). *Map3k8* deficiency in either nHCs or HCs in surviving mice increased pulmonary edema (Fig. 4E) and collagen expression (Fig. 4G, 4H) in comparison with wt/wt controls, indicating that *Map3k8* expression from both nHCs and HCs exert some protective role in disease pathogenesis. Noteworthy, the accumulation of inflammatory cells in the BALF was more pronounced when *Map3k8* expression was missing from the HCs (Fig. 4F). Accordingly, and in agreement with the systemic manifestations (Fig. 4C, 4D), *Map3k8* deficiency in HCs was sufficient to exacerbate collagen deposition and to distort lung architecture (Fig. 4I), suggesting a main protective role in pulmonary fibrosis for *Map3k8* expression in HCs.

Because the deletion of *Map3k8* in HCs had a higher effect in the number of inflammatory cells in the BALFs of BLM-

challenged mice (Fig. 4F), we next performed basic (non-exhaustive) FACS analysis of immune cells in the BALFs (Fig. 5A). *Map3k8* deletion had minor effects in neutrophilic infiltration (Fig. 5B); accordingly, *Map3k8*^{-/-} mice had minor and nonconsistent effects in modeled acute lung disorders, such as LPS-induced lung injury and ventilation-induced lung injury (data not shown). However, *Map3k8* deletion in either of the cellular reservoirs promoted the accumulation of macrophages, which was found to be more pronounced upon deletion in HCs (Fig. 5B). Moreover, *Map3k8* deletion in HCs, but not in nHCs, further promoted lymphocyte accumulation (Fig. 5C).

Given the suggested role(s) of *Map3k8* in macrophage responses, its reduced expression in fibrotic lungs (Figs. 1, 2) and IPF macrophages (Supplemental Fig. 1), and the finding that *Map3k8* deficiency in HCs exacerbates BLM-induced pulmonary inflammation and fibrosis (Figs. 4, 5), we next genetically deleted *Map3k8* specifically in macrophages (and granulocytes) by mating the conditional (floxed; ^{f/f}) knockout mouse for *Map3k8* (*Map3k8*^{f/f}) (12) with a transgenic mouse strain expressing the Cre recombinase under the control of the *LysM* promoter (*TgLysM-Cre*) (11). *TgLysM-Cre* has been reported with an 80–95% deletion efficiency in macrophages (11, 26), whereas *LysM*-mediated Cre expression, per se, has been previously reported not to have any effects on lung development and architecture or BLM-induced pulmonary fibrosis (26). BLM was then administered to *LysM-Map3k8*^{-/-} mice and littermate controls and disease severity was assessed with the standardized readout assays at the peak of the

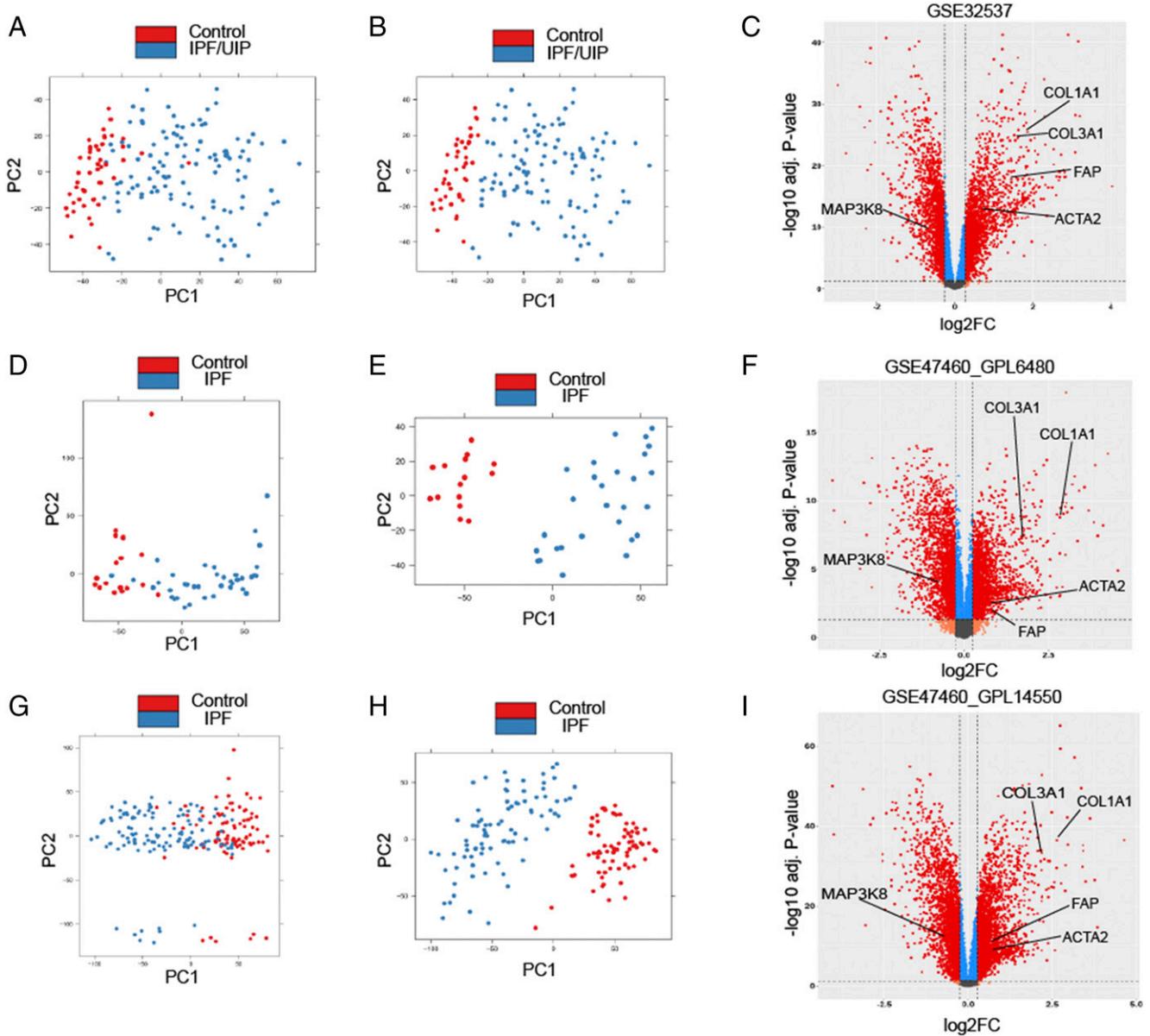


FIGURE 2. Decreased *MAP3K8* mRNA expression in the fibrotic lung tissue of IPF patients. Reanalysis of two microarray datasets (GSE32537 and GSE47460) interrogating differential expression in the lung tissue of 119, 122, and 38 IPF patients in comparison with 50, 91, and 17 controls, respectively. Raw data were background corrected and rma normalized, and outliers were removed based on PCA plots (A–D, G, and H before and after curation, respectively). Differential expression analysis of gene level–summarized intensity values is presented using volcano plots (C, F, and I). A gene was considered statistically significantly differentially expressed when having a fold change (FC) score >1.2 or <0.07 at a false discover rate (FDR) of 5% (in Red). A modest but highly statistically significant decrease of *MAP3K8* mRNA levels was detected in IPF patients in all three cohorts/datasets as indicated in the text; the expression of well-established fibrotic genes (*COL1A1*, *COL3A1*, *ACTA2*, and *FAP*) was found upregulated in the same samples.

disease. Disease development in *LysMMap3k8^{-/-}* mice was found exacerbated in all readout assays, including respiratory functions (Fig. 6). Furthermore, FACS analysis of inflammatory cells in the BALF in the absence of *Map3k8* expression from macrophages (Fig. 7) phenocopied the HC deletion (Fig. 5), indicating a major protective role for macrophage (*LysM⁺*) *Map3k8* expression in the disease pathogenesis as well as confirming the seminal contribution of macrophages in pulmonary fibrosis.

Map3k8 regulates the cyclooxygenase-2–*PGE₂* axis, which exerts anti-inflammatory and antifibrotic effects in the lung

Among the different cellular pathways affected by *Map3k8* that could possibly play a protective role in pulmonary fibrosis, *Map3k8* has been reported to regulate cyclooxygenase-2 (*Cox-2*)–mediated *PGE₂* expression from arachidonic acid (AA) (Fig. 8A) (12, 18). *PGE₂* is a

bioactive eicosanoid that is considered as a proinflammatory mediator; on the contrary, in the lung, *PGE₂* has an established role in limiting fibrotic processes (27); however, its local regulation remains relatively unexplored and its cellular origin uncertain.

To examine if *Cox-2* mRNA levels are affected by the genetic deletion of *Map3k8*, we performed Q-RT-PCR analysis in lung tissue samples from genetically modified mice at the peak of the disease post-BLM administration. Ubiquitous genetic deletion of *Map3k8* (Fig. 8B) was found to downregulate *Cox-2* mRNA levels (Fig. 8C). A similar expression profile was also observed upon the genetic deletion of *Map3k8* in nHCs and HCs in chimeric mice (Supplemental Fig. 2A, 2B) as well as in macrophages (Supplemental Fig. 2C, 2D), thus confirming 1) *Map3k8* expression from both nHCs and HCs, 2) efficient *Map3k8* targeting, as well as 3) the regulatory role of *Map3k8* in *Cox-2* expression.

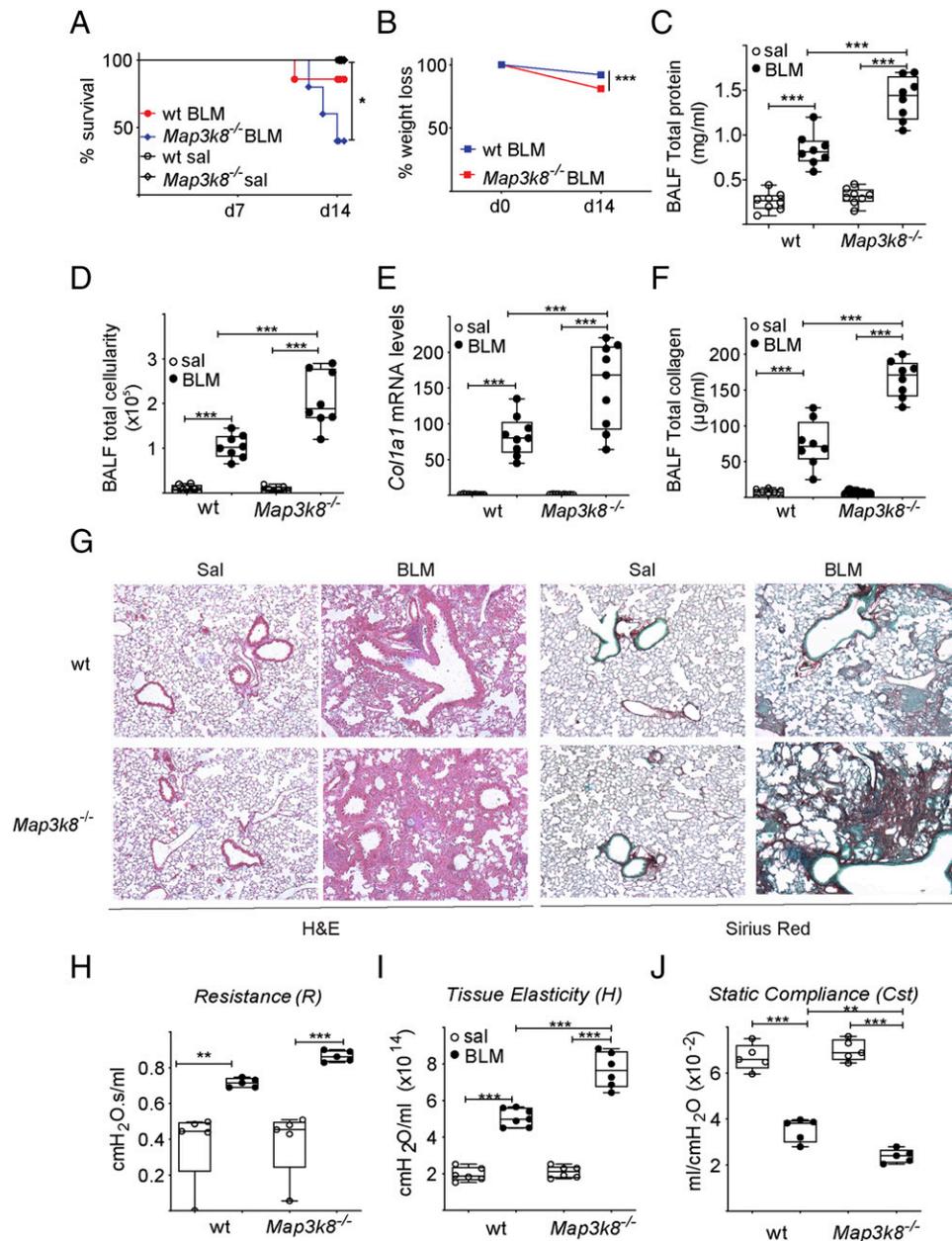


FIGURE 3. Ubiquitous genetic deletion of *Map3k8* exacerbates pulmonary fibrosis. BLM was administered to *Map3k8*-deficient mice and wt littermates, and disease development was assessed with standardized assays 14 d post-BLM administration. **(A)** Kaplan Meyer survival curve. **(B)** Weight loss post-BLM administration. **(C)** Total protein concentration in BALFs, as determined with the Bradford assay. **(D)** Inflammatory cell numbers in BALFs, as counted with a hemacytometer. **(E)** *Col1a1* mRNA levels in whole-lung tissue were determined with Q-RT-PCR analysis; values were normalized to the expression of B2M. **(F)** Soluble collagen in BALFs was determined with the Sirius Red assay. **(G)** Representative H&E- and Sirius Red-stained sections of murine lungs of the indicated genotypes (original magnification $\times 10$). **(H–J)** Indicated respiratory functions were measured with FlexiVent. Statistical significance was assessed with two-way ANOVA with Bonferroni post hoc correction. Data are presented as means \pm SEM ($n = 5$ – 10 ; cumulative results from two separate experiments). * $p = 0.05$, ** $p = 0.01$, *** $p = 0.001$.

Accordingly, AA metabolism was found to be disturbed in *Map3k8*^{-/-} mice upon BLM-induced pulmonary fibrosis that presented with increased levels of AA (Fig. 8D) and reduced levels of PGE₂ in their BALFs (Fig. 5E–G), as deduced, given its extreme instability from its metabolites, according to an established and recently published procedure (18). Similar results were obtained in chimeric mice (Supplemental Fig. 3), thus establishing Map3k8 as a major regulator of PGE₂ production in the lungs and suggesting HCs as the major PGE₂ producers in the lung.

To examine if the decreased PGE₂ levels upon *Map3k8* deletion contribute to the observed exacerbated fibrotic phenotype, dmpGE₂, a stable analogue of endogenous PGE₂, was administered (i.p.; 10 µg/kg;

twice daily) to *Map3k8*^{-/-} and wt littermate mice undergoing BLM-induced pulmonary fibrosis development. dmpGE₂ administration restored the exacerbated fibrotic responses in *Map3k8*^{-/-} mice, as indicated from all readout assays (Fig. 9), as well as decreased the severity of the BLM-induced disease in wt mice (Fig. 9). Noteworthy, PGE₂ administration also attenuated inflammatory influx (Fig. 10), integral to fibrosis development in this animal BLM model.

Discussion

MAP3K8 mRNA was found downregulated in both IPF patients and the corresponding animal model, whereas public scRNAseq

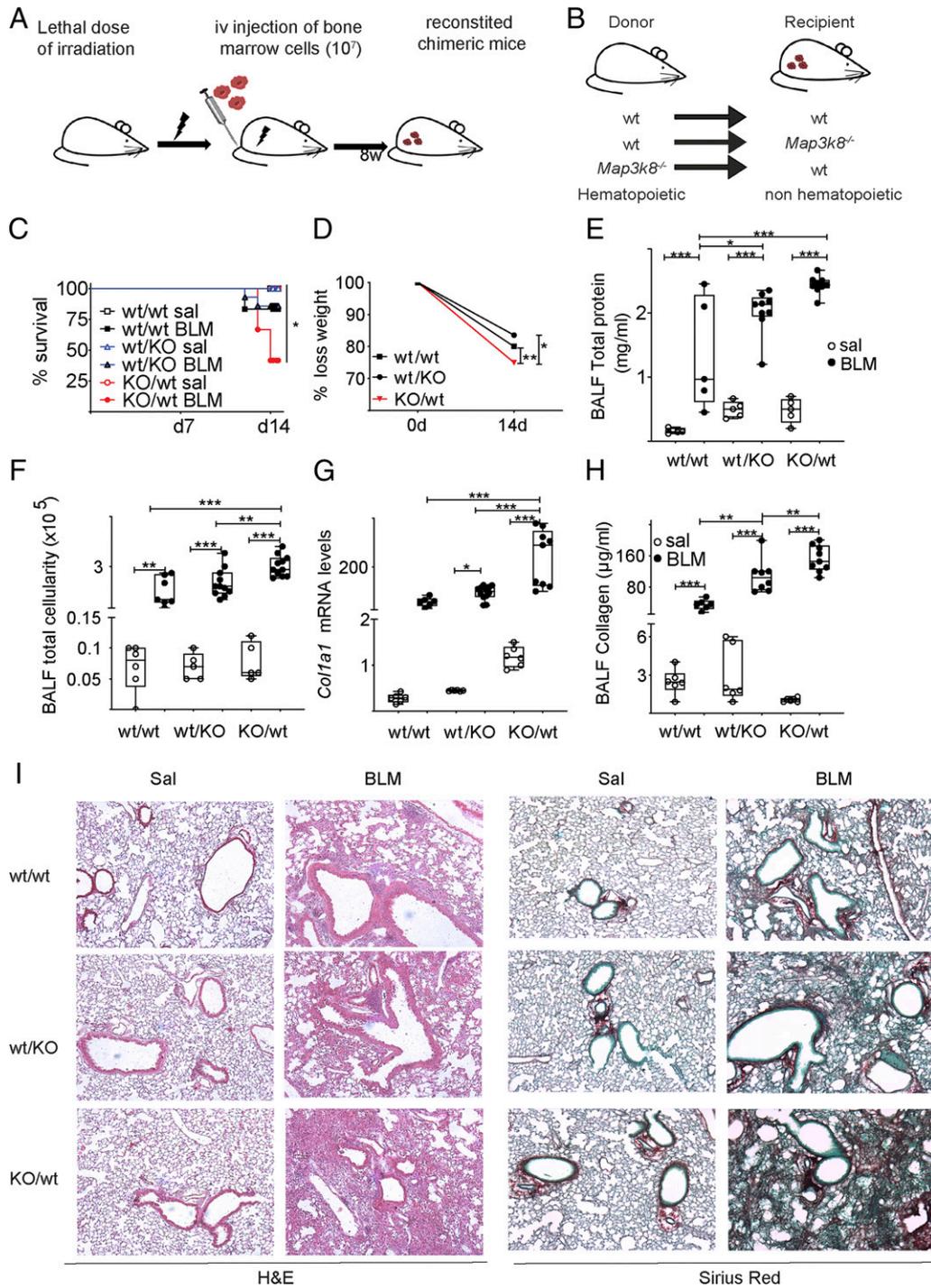
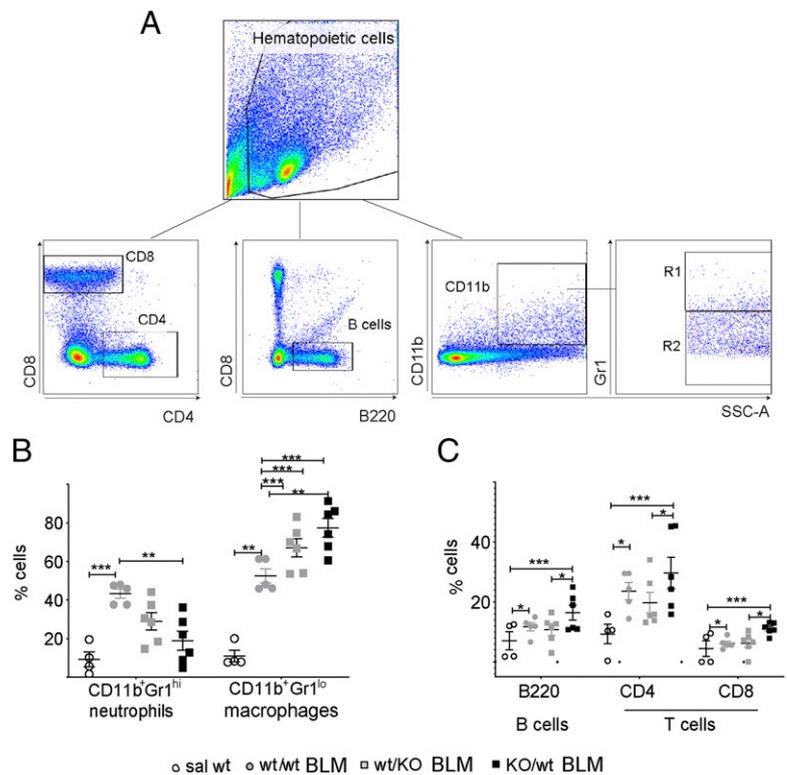


FIGURE 4. Genetic deletion of *Map3k8* from either nHCs or HCs exacerbates pulmonary inflammation and fibrosis. BLM was administered to chimeric mice bearing HCs with the genetic background of the donor as indicated: wt or knockout (KO) (*Map3k8*^{-/-}) and disease development was assessed with standardized assays 14 d post-BLM. **(A)** Schematic representation of bone marrow transfers. **(B)** Obtained chimeric mice from bone marrow transfers. **(C)** Kaplan Meyer survival post-BLM administration. **(D)** Weight loss post-BLM administration. **(E)** Total protein concentration in BALFs, as determined with the Bradford assay. **(F)** Inflammatory cell numbers in BALFs, as counted with a hemacytometer. **(G)** *Col1a1* mRNA levels in whole-lung tissue were determined with Q-RT-PCR analysis; values were normalized to the expression of B2M. **(H)** Soluble collagen in BALFs was determined with the Sirius Red assay. **(I)** Representative H&E- and Sirius Red-stained sections of murine lungs of the indicated genotypes (original magnification $\times 10$). Statistical significance was assessed with two-way ANOVA with Bonferroni post hoc correction. Data are presented as means \pm SEM ($n = 5$ –12; cumulative results from two separate experiments). * $p = 0.05$, ** $p = 0.01$, *** $p = 0.001$.

data mining suggested that *MAP3K8* is expressed in both nHCs and HCs in the human lung. Genetic deletion of *Map3k8* in mice exacerbated BLM-induced pulmonary fibrosis, whereas bone marrow transfer experiments indicated that although *Map3k8* expression in both nHCs and HCs exert some protective role in

disease pathogenesis, *Map3k8* in HCs has a more dominant role. Moreover, macrophage-specific deletion of *Map3k8* was shown to be sufficient to exacerbate disease severity, thus confirming a major role for macrophages in fibrotic responses in the lung and a role for *Map3k8* in the homeostasis of their effector functions.

FIGURE 5. *Map3k8* deficiency in either nHCs or HCs promotes the accumulation of inflammatory cells in the BALF upon BLM-induced pulmonary fibrosis. **(A)** Gating strategy of FACS analysis. **(B and C)** FACS analysis of the indicated cell types with the indicated cell markers. Data are presented as mean \pm SEM. Statistical significance was assessed with two-way ANOVA with Bonferroni post hoc correction ($n = 5-7$; cumulative results from two separate experiments are shown). * $p = 0.05$, ** $p = 0.01$, *** $p = 0.001$.



Map3k8 deficiency was further shown to decrease *Cox-2* mRNA expression, followed by a decrease in PGE_2 expression in the lung, whereas exogenous administration of (dm) PGE_2 reversed the exacerbated fibrotic response of *Map3k8*^{-/-} mice and decreased disease severity in wt mice.

The expression and/or activation of *MAP3K8* has been shown to be modulated by multiple inflammatory mediators through their cognate receptors, such as TLRs, CD40, TNFR1, IL-1R, as well as different GPCRs. At steady state, *MAP3K8* is inactive through the association with NF- κ B p105 and A20-binding inhibitor of NF- κ B 2 (ABIN2). Activation of I κ B kinase (IKK β) leads to the phosphorylation of p105 and its subsequent ubiquitination and proteasomal degradation, thus releasing *MAP3K8*, which is then phosphorylated, in turn activating its downstream targets and thus participating in the regulation of inflammatory responses (7). However, and given the central role of *Map3k8* in regulating inflammatory and homeostatic responses, additional levels of regulation of the *Map3k8* levels and effector functions have been reported at the genetic and epigenetic level. *MAP3K8* mRNA expression has been found decreased in intestinal myofibroblasts isolated from the inflamed ileum of inflammatory bowel disease patients (12). Decreased *MAP3K8* mRNA expression was also detected in the lung tissue of lung cancer patients correlating with poor survival, suggested to be imposed through miRNA-370 which targets *MAP3K8* transcripts for degradation (9). As shown in this study, decreased *MAP3K8* mRNA levels were detected in the lungs of mice upon BLM-induced pulmonary fibrosis (Fig. 1) and of IPF patients (Fig. 2), indicating transcriptional or posttranscriptional downregulation of *Map3k8* expression upon fibrinogenesis in the lung, notwithstanding additional means of regulation via protein-protein interactions and/or phosphorylation cascades.

Irrespectively of the regulatory mode of *Map3k8* expression, low *Map3k8* levels have been associated with reduced *Cox-2* and PGE_2 levels in different cell types and pathophysiological

situations, including inflamed adipocytes (28), intestinal myofibroblasts (12), and activated macrophages (29); however, no such correlation has been reported in the context of pulmonary fibrosis. As shown in this study, reduced *Cox-2* and PGE_2 levels always accompanied *Map3k8* genetic deficiency and the associated exacerbated fibrotic phenotype (Figs. 8, Supplemental Fig. 2), which could be reverted by the administration of exogenous PGE_2 (Figs. 9, 10). PGE_2 administration also prevented disease development in wt mice, as previously reported (30, 31), whereas no gross pathologic effects were observed (data not shown) (30, 31).

Therefore, *Map3k8* plays a central role in the regulation of local PGE_2 production in the lung. However, additional means of regulation of PGE_2 production have been proposed, via soluble mediators and epigenetic changes (reviewed in Ref. 27). Moreover, little is known on the relative contribution of different cell types in the overall PGE_2 production, although the current dogma, based mostly on in vitro experiments, suggests alveolar epithelial cells and fibroblasts as the main producers in the lung (27). However, as shown in this study in vivo (Fig. 4), both nHCs and HCs produce PGE_2 , whereas the overall contribution of HCs (likely macrophages) in pulmonary levels was found higher in these experimental settings.

Map3k8^{-/-} mice have been previously reported to be protected from LPS/D-gal-induced endotoxic shock (32), attributed to defective TNF and chemokine receptors expression from macrophages (32, 33). *Map3k8*^{-/-} mice were also found protected from experimental encephalomyelitis, attributed to defective IL-17 signaling (34), as well as from Con A (ConA)-induced, T cell-dependent and TNF-mediated liver inflammation and injury, attributed to defective NKT effector functions (35). However, and on the contrary, *Map3k8* ubiquitous genetic deficiency has been reported to elevate inflammation and to exacerbate fibrosis in the small intestine, liver, and lung following *Schistosoma mansoni* infection (36), as well as to promote intestinal inflammation and

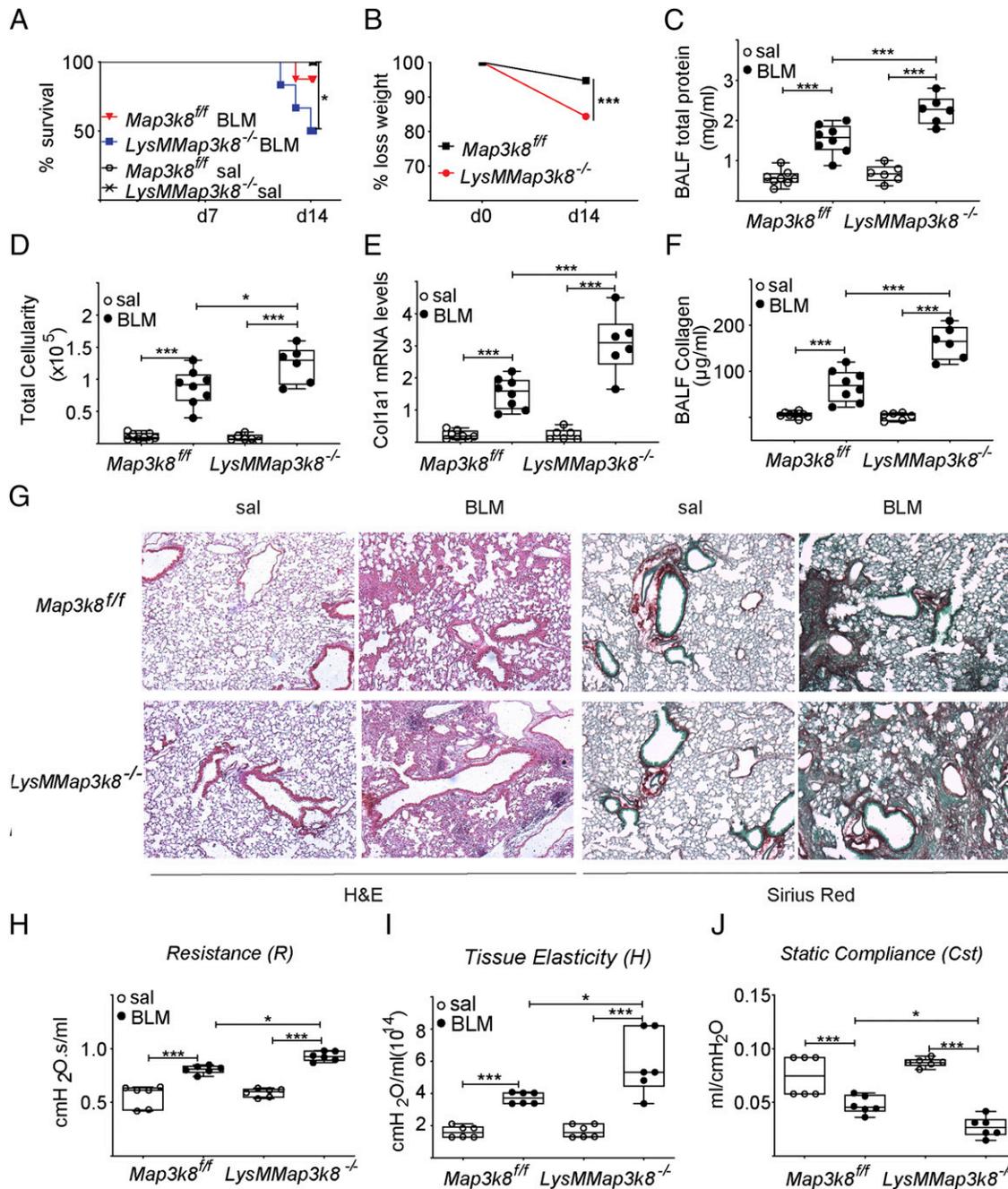
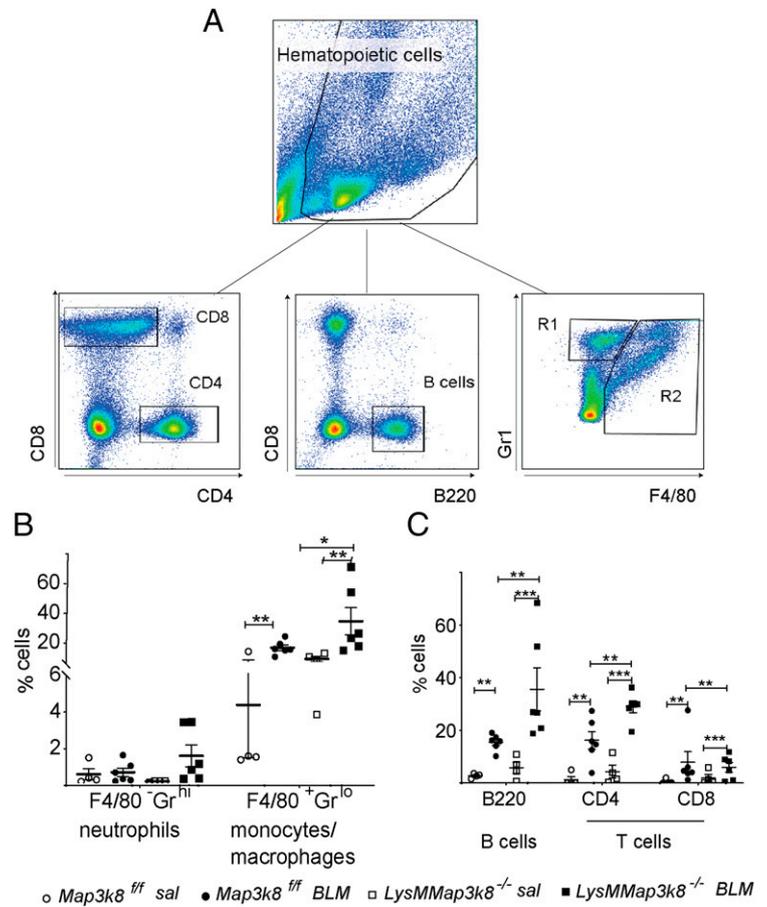


FIGURE 6. Genetic deletion of *Map3k8* from macrophages (*LysM⁺* cells) exacerbates pulmonary inflammation and fibrosis. BLM was administered to *LysMMap3k8^{-/-}* mice lacking *Map3k8* in macrophages (and granulocytes) and wt littermates, and disease development was assessed with standardized assays 14 d post-BLM. **(A)** Kaplan Meyer survival and **(B)** weight loss post-BLM administration. **(C)** Total protein concentration in BALFs, as determined with the Bradford assay. **(D)** Inflammatory cell numbers in BALFs, as counted with a hemacytometer. **(E)** *Col1a1* mRNA levels in whole-lung tissue were determined with Q-RT-PCR analysis; values were normalized to the expression of *B2M*. **(F)** Soluble collagen in BALFs was determined with the Sirius Red assay. **(G)** Representative H&E- and Sirius Red-stained sections of murine lungs of the indicated genotypes (original magnification $\times 10$). **(H–J)** Indicated respiratory functions were measured with flexiVent. Statistical significance was assessed with two-way ANOVA with Bonferroni post hoc correction. Data are presented as means \pm SEM ($n = 5–8$; cumulative results from two separate experiments). * $p = 0.05$, *** $p = 0.001$.

tumorigenesis (12, 37). In agreement, in this report, genetic deletion of *Map3k8* was shown to exacerbate BLM-induced pulmonary fibrosis (Fig. 3). A similar protective role was previously shown for urethane-induced lung carcinogenesis (9), thus extending the mechanistic similarities between IPF and lung cancer (10) and suggesting a major homeostatic role for *Map3k8* in lung pathophysiology and fibrotic responses that may underlie carcinogenesis.

Map3k8 is expressed from both nHCs and HCs in the lung (Supplemental Fig. 1) (7), and genetic deficiency in either cellular reservoir was shown to exacerbate the modeled disease (Fig. 4), suggesting that MAP3K8-regulated homeostatic mechanisms are active in both compartments. Although MAP3K8 is thought to be expressed mainly in HCs, it can also be detected in other cells (7), including adipocytes (28) and fibroblasts (Supplemental Fig. 1) (12). Decreased *MAP3K8* expression was visualized in IPF

FIGURE 7. *Map3k8* deficiency in LysM⁺ cells promotes the accumulation of inflammatory cells in the BALF upon BLM-induced pulmonary fibrosis. **(A)** Gating strategy of FACS analysis. **(B and C)** FACS analysis of the indicated cell types with the indicated cell markers. Data are presented as mean \pm SEM. Statistical significance was assessed with two-way ANOVA with Bonferroni post hoc correction ($n = 5-7$; cumulative results from two separate experiments are shown). * $p = 0.05$, ** $p = 0.01$, *** $p = 0.001$.

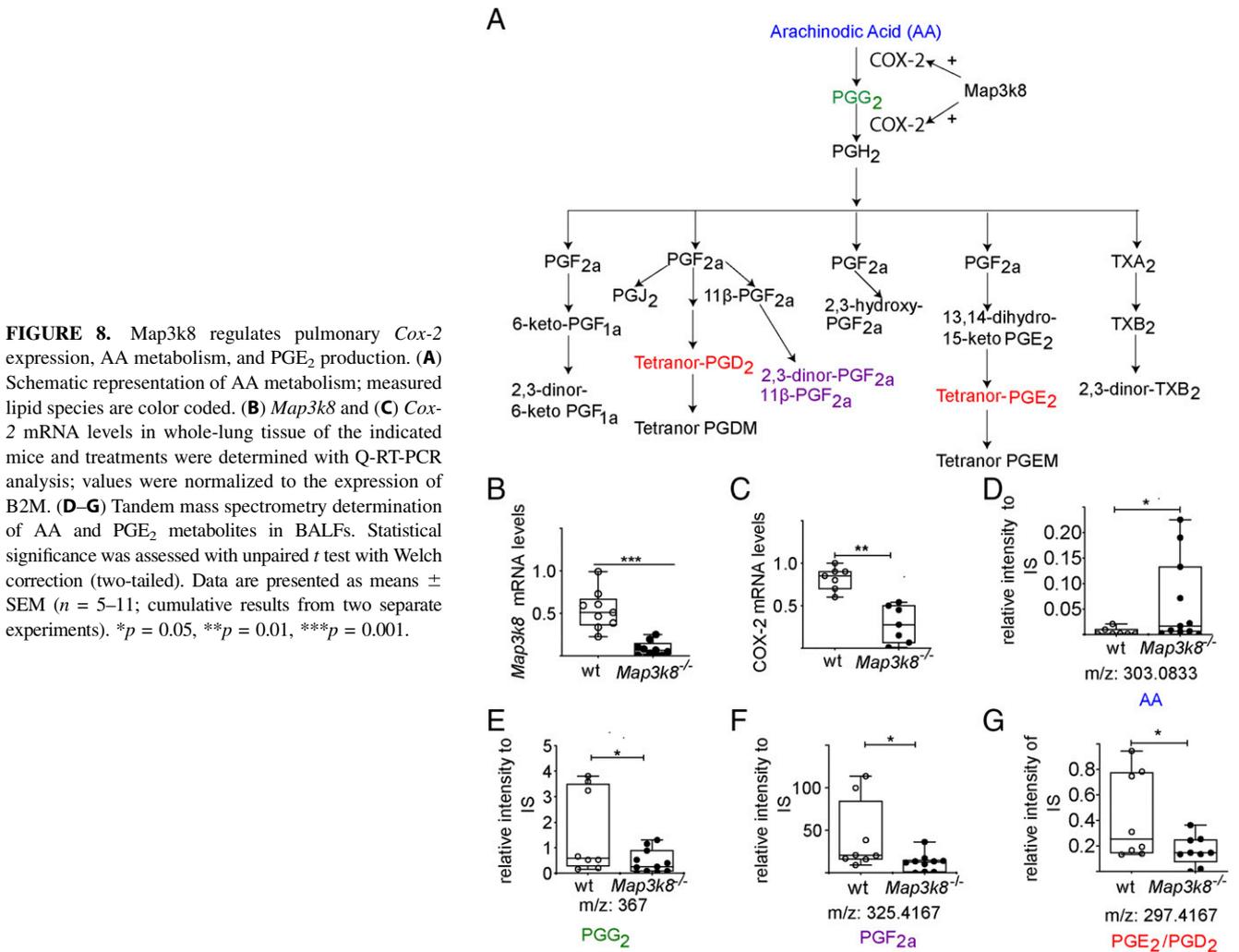


(α SMA/ACTA2⁺) pulmonary myofibroblasts (Supplemental Fig. 1B, 1C), whose accumulation is a major hallmark of pulmonary fibrosis (1, 2). Decreased *MAP3K8* expression was previously shown for inflammatory bowel disease intestinal myofibroblasts (12), correlating with reduced COX-2 expression and PGE₂ production, suggesting similar effects in IPF myofibroblasts. Accordingly, lung fibroblasts isolated from IPF patients have been suggested to express reduced levels of COX-2 and PGE₂ (38, 39). Moreover, alveolar epithelial cells have been proposed to produce PGE₂, via Cox-2, which target adjacent fibroblasts (40). In turn, PGE₂ has been shown to decrease fibroblast differentiation and collagen production (reviewed in Ref. 27) and to promote their apoptosis (41), likely through the EP2 receptor and disruption of calcium signaling (42). Therefore, the observed exacerbated fibrotic response of chimeric mice lacking *Map3k8* in nHCs cells (Fig. 5) can be partly attributed to the attenuation of the anti-fibrotic effects of PGE₂ on lung fibroblasts.

However, the bone marrow transfer experiments indicated, for the first time (to our knowledge) on a quantitative basis in vivo with tandem mass spectrometry, that HCs are the major source of *Map3k8*-regulated, Cox-2-mediated, PGE₂ production in the lung in modeled pulmonary fibrosis (Figs. 8, Supplemental Fig. 3). Moreover, *Map3k8* deletion in HCs, phenocopied in the macrophage-specific deletion, was sufficient for disease exacerbation (Figs. 4, 6), corroborating the important role of *Map3k8* in the regulation of macrophage responses, as well as the role of pulmonary macrophages, per se, in fibrosis development in the lung. Macrophages are well recognized as essential players in IPF pathogenesis, as they comprise the major immune cell type populating the lungs of IPF patients, whereas depletion of circulating monocytes, via genetic or pharmacologic means, attenuates

fibrosis severity in animal models (43–47). Beyond their well-established roles in apoptotic cell clearance and the production of profibrotic mediators such as TGF β and IL-13, macrophages secrete numerous cytokines and chemokines thus modulating the immune response, as well as matrix metalloproteinases responsible for extracellular matrix remodeling and resolution. However, their mode of action remains controversial because they are highly heterogeneous and exhibit remarkable plasticity (48, 49).

Pulmonary macrophages can be grouped into two broad subsets based on their anatomic location: alveolar macrophages, which line the surface of alveoli, and interstitial macrophages (IMs) localized between the alveolar epithelium and the vascular endothelium. Furthermore, mouse alveolar macrophages include tissue-resident cells, which are long-lived, self-renewing cells that arise from fetal progenitors (50, 51), as well as monocyte-derived alveolar macrophages (Mo-AMs) (52, 53). Mo-AMs originate postnatally from circulating monocytes, are recruited via a CCL2/CCR2 axis, exhibit a proinflammatory expression profile, and have been suggested to be the main macrophage subgroup driving pulmonary fibrosis in mice (54, 55). IMs also originate from monocytes, and some from the yolk sac are reprogrammed epigenetically by the local microenvironment, although it has been suggested that they may serve as an obligatory intermediate between blood monocytes and alveolar macrophages (44, 56). Three distinct populations of IMs have been suggested (IM1-3) based on surface markers exhibiting differential turnover rates (57). The advent of scRNAseq has revealed even greater macrophage diversity, and several novel macrophage subsets have been suggested, defined by the expression of different chemokines and cell-surface markers such as M-CSF/M-CSFR, SPP1/MERTK, CX3CR1, Ly6C^{hi}, CD171,



Fra-2, CD300c2, and Lrp5. However, a consensus has not been reached because data analysis of scRNAseq data and cluster identification have not yet developed standard operational procedures and commonly used cell-specific markers. Given the differential roles that have been suggested for the different macrophage subsets and to possibly identify in which macrophage subtype Map3k8 is found downregulated, we performed reanalysis of the publicly available scRNAseq dataset GSE122960 (23) (Supplemental Fig. 1). *MAP3K8* was found downregulated in a profibrotic macrophage subpopulation characterized by elevated expression of *SPP1*, likely corresponding to the previously reported macrophage subset SPP^{hi} in humans (58), as well as to the Mo-AMs in mice (54), reported to drive the pathogenesis of the disease (54, 58). However, de novo scRNAseq of both lung and BALF cells in the absence of *Map3k8* will be further required to fully appreciate the role of Map3k8 in macrophage subsets and to obtain additional mechanistic insights.

Increasing further heterogeneity and complexity, macrophages can also get polarized, depending on local stimuli, toward two highly dynamic and overlapping states of activation, classically activated (by IFN- γ and TNF) M1 macrophages or alternatively activated (by IL-4, IL-13, etc.) M2 macrophages. M1 macrophages, which secrete TNF, IL1b, and IL-6, are considered proinflammatory, whereas M2 macrophages (that can be further subclassified into a–d), which secrete IL-10 and TGF among others, are thought to suppress inflammation while promoting

fibroproliferation and uncontrolled repair. *Map3k8* deficiency has been suggested to promote macrophage polarization to M2, following *S. mansoni* infection (36), suggesting yet another profibrotic mechanism that could be regulated by Map3k8 in macrophages. The promotion of macrophage M2 polarization upon reduced levels of Map3k8 in IPF macrophages could influence not only macrophage wound healing effector functions and fibrosis, but also T cell physiology and functions. As shown in this study, *Map3k8* genetic deletion in HCs (Fig. 3) or specifically in macrophages (Fig. 4) increased inflammatory cells in BALFs, including T cells. Moreover, *Map3k8* ubiquitous genetic deletion has been shown to promote Th2 polarization of the T cell response in the lung (8, 36), whereas IPF is considered as a type II disease and type 2 cytokines, such as IL-4 and IL-13, are elevated in IPF (59). T cell functions can be also modulated by yet another monocytic subtype, namely circulating myeloid-derived suppressor cells. Increased numbers of myeloid-derived suppressor cells have been found in both IPF patients and the BLM model, suggested to modulate IPF progression by orchestrating immunosuppressive and profibrotic networks (60, 61). Immunosuppressive signals have been also suggested to emanate from a sessile sub population of alveolar macrophages via direct communication, through Cx43, with alveolar epithelial cells (62). Moreover, macrophages have been suggested to direct the metabolism and homeostasis of adjacent cells (63) such as fibroblasts whose activation includes glycolytic reprogramming (64).

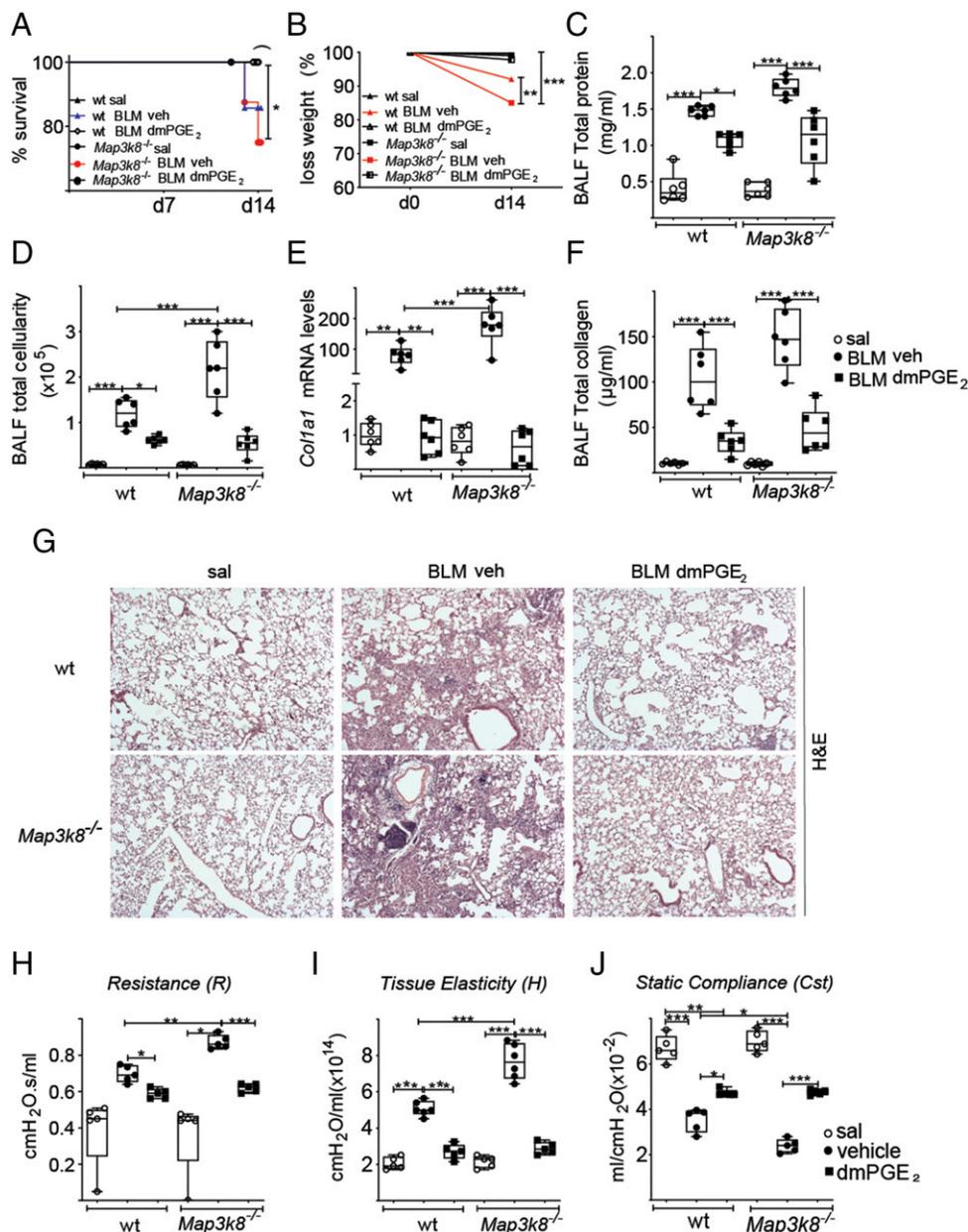


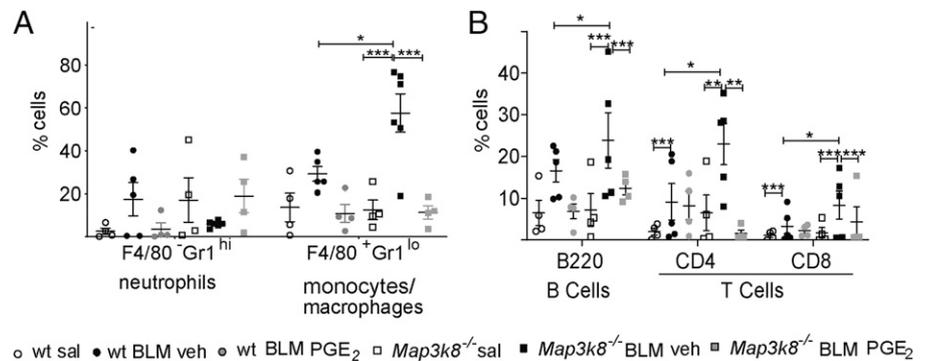
FIGURE 9. PGE₂ restores the profibrotic effects of *Map3k8* deficiency and attenuates pulmonary inflammation and fibrosis. dmPGE₂ was administered (i.p.; 10 µg/kg; twice daily) to *Map3k8*^{-/-} and wt littermate mice undergoing BLM-induced pulmonary fibrosis, and disease development was assessed with standardized assays 14 d post-BLM. **(A)** Kaplan Meyer survival and **(B)** weight loss post-BLM administration. **(C)** Total protein concentration in BALFs, as determined with the Bradford assay. **(D)** Inflammatory cell numbers in BALFs, as counted with a hemacytometer. **(E)** *Col1a1* mRNA levels in whole-lung tissue were determined with Q-RT-PCR analysis; values were normalized to the expression of B2M. **(F)** Soluble collagen in BALFs was determined with the Sirius Red assay. **(G)** Representative H&E-stained sections of murine lungs of the indicated genotypes and treatments (original magnification ×10). **(H–J)** Indicated respiratory functions were measured with flexiVent. Statistical significance was assessed with two-way ANOVA with Bonferroni post hoc correction. Data are presented as means ± SEM (*n* = 6). **p* = 0.05, ***p* = 0.01, ****p* = 0.001.

Notwithstanding the *Map3k8* regulation of PGE₂ production, *Map3k8* has been previously suggested to modulate, via MAPK pathways, the expression of various cytokines and chemokines (65). More specifically, *Map3k8* genetic deletion has been reported to decrease the levels of TNF and IL-1b (32), IL-6, CXCL2 (MIP-2), and CCL2 (MCP-1) on nonmyeloid cells (66), as well as CCR1, CCR2, and CCR5 on macrophages (33). All of these proinflammatory mediators have been found deregulated in pulmonary fibrosis (48, 49, 65), including the BLM-induced disease (data not shown) (15, 26); among them, CCL2 has been shown to be critical for macrophage recruitment (43). mRNA levels of TNF, IL-1b, IL-6, CXCL2, and

CCL2 were found variably decreased in the lung tissue in the absence of *Map3k8* at the peak of the BLM-induced fibrosis (data not shown). Therefore, and although RT-PCR in whole-lung tissue at the peak of the disease is not the optimal methodology for this purpose, the effect of *Map3k8* deletion on the levels of proinflammatory mediators cannot account for the increased fibrosis post-BLM administration in *Map3k8* null mice in these settings.

Further studies, employing inducible deletion of *Map3k8* and/or the use of other fibrotic animal models that bypass inflammation (i.e., adenoviral delivery of TGFb), will be required to dissect the effects of *Map3k8*/Cox-2/PGE₂ axis on fibrosis, per se. Moreover,

FIGURE 10. PGE₂ attenuates inflammatory cell influx in the lung. dmPGE₂ was administered (i.p.; 10 μg/kg; twice daily) to *Map3k8*^{-/-} and wt littermate mice undergoing BLM-induced pulmonary fibrosis development. (A and B) FACS analysis of the indicated cell types in BALFs with the indicated cell markers. Gating strategy as in Fig. 7. Data are presented as mean ± SEM. Statistical significance was assessed with two-way ANOVA with Bonferroni multiple comparisons test (*n* = 3–6). **p* = 0.05, ***p* = 0.01, ****p* = 0.001.



further lipidomic analysis in IPF and its animal model will be necessary to fully appreciate the pleiotropic effects of various lipids (leukotrienes, thromboxanes, and PGs, as well as lysophospholipids) on pulmonary inflammation and fibrosis.

In conclusion, *Map3k8* was shown to have a major role in healthy lung homeostasis, regulating, among others, the local production of PGE₂ and its antifibrotic effects. The development of fibrosis, in both humans and mice, entails a transcriptional downregulation of *Map3k8* expression in different pulmonary cells, and especially macrophages. Down regulation of *Map3k8* levels results, possibly among others, in the abrogation of Cox-2-mediated PGE₂ production and the alleviation of its lung specific antifibrotic effects.

Acknowledgments

We thank G. Kollias (Biomedical Sciences Research Center Fleming, National and Kapodistrian University of Athens) for the *Map3k8*^{fl/fl} mice and M. Roulis (Yale University) for critical reading of the manuscript.

Disclosures

The authors have no financial conflicts of interest.

References

- Mora, A. L., M. Rojas, A. Pardo, and M. Selman. 2017. Emerging therapies for idiopathic pulmonary fibrosis, a progressive age-related disease. [Published erratum appears in 2017 *Nat. Rev. Drug Discov.* 16: 810.] *Nat. Rev. Drug Discov.* 16: 755–772.
- Martinez, F. J., H. R. Collard, A. Pardo, G. Raghu, L. Richeldi, M. Selman, J. J. Swigris, H. Taniguchi, and A. U. Wells. 2017. Idiopathic pulmonary fibrosis. *Nat. Rev. Dis. Primers* 3: 17074.
- Noble, P. W., C. Albera, W. Z. Bradford, U. Costabel, M. K. Glassberg, D. Kardatzke, T. E. King, Jr., L. Lancaster, S. A. Sahn, J. Szwarcberg, et al; CAPACITY Study Group. 2011. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet* 377: 1760–1769.
- Richeldi, L., R. M. du Bois, G. Raghu, A. Azuma, K. K. Brown, U. Costabel, V. Cottin, K. R. Flaherty, D. M. Hansell, Y. Inoue, et al; INPULSIS Trial Investigators. 2014. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. [Published erratum appears in 2015 *N. Engl. J. Med.* 373: 782.] *N. Engl. J. Med.* 370: 2071–2082.
- Griminger, F., A. Günther, and C. Vancheri. 2015. The role of tyrosine kinases in the pathogenesis of idiopathic pulmonary fibrosis. *Eur. Respir. J.* 45: 1426–1433.
- Cargnello, M., and P. P. Roux. 2011. Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases. *Microbiol. Mol. Biol. Rev.* 75: 50–83.
- Vougioukalaki, M., D. C. Kanellis, K. Gkouskou, and A. G. Eliopoulos. 2011. Tpl2 kinase signal transduction in inflammation and cancer. *Cancer Lett.* 304: 80–89.
- Watford, W. T., C. C. Wang, C. Tsatsanis, L. A. Mielke, A. G. Eliopoulos, C. Daskalakis, N. Charles, S. Odom, J. Rivera, J. O’Shea, and P. N. Tschlis. 2010. Ablation of tumor progression locus 2 promotes a type 2 Th cell response in Ovalbumin-immunized mice. *J. Immunol.* 184: 105–113.
- Gkirtzimanaki, K., K. K. Gkouskou, U. Oleksiewicz, G. Nikolaidis, D. Vyrila, M. Lontos, V. Pelekanou, D. C. Kanellis, K. Evangelou, E. N. Stathopoulos, et al. 2013. TPL2 kinase is a suppressor of lung carcinogenesis. *Proc. Natl. Acad. Sci. USA* 110: E1470–E1479.
- Karampitsakos, T., V. Tzilas, R. Tringidou, P. Steiropoulos, V. Aidinis, S. A. Papiiris, D. Bouros, and A. Tzouveleakis. 2017. Lung cancer in patients with idiopathic pulmonary fibrosis. *Pulm. Pharmacol. Ther.* 45: 1–10.
- Clausen, B. E., C. Burkhardt, W. Reith, R. Renkawitz, and I. Förster. 1999. Conditional gene targeting in macrophages and granulocytes using LysMcre mice. *Transgenic Res.* 8: 265–277.
- Roulis, M., C. Nikolauou, E. Kotsaki, E. Kaffe, N. Karagianni, V. Koliariaki, K. Salpea, J. Ragoussis, V. Aidinis, E. Martini, et al. 2014. Intestinal myofibroblast-specific Tpl2-Cox-2-PGE₂ pathway links innate sensing to epithelial homeostasis. *Proc. Natl. Acad. Sci. USA* 111: E4658–E4667.
- Barbayaanni, I., I. Ninou, A. Tzouveleakis, and V. Aidinis. 2018. Bleomycin revisited: a direct comparison of the intratracheal micro-spraying and the oropharyngeal aspiration routes of bleomycin administration in mice. *Front. Med. (Lausanne)* 5: 269.
- Jenkins, R. G., B. B. Moore, R. C. Chambers, O. Eickelberg, M. Königshoff, M. Kolb, G. J. Laurent, C. B. Nanthakumar, M. A. Olman, A. Pardo, et al; ATS Assembly on Respiratory Cell and Molecular Biology. 2017. An official American Thoracic Society workshop report: use of animal models for the preclinical assessment of potential therapies for pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* 56: 667–679.
- Mouratis, M. A., and V. Aidinis. 2011. Modeling pulmonary fibrosis with bleomycin. *Curr. Opin. Pulm. Med.* 17: 355–361.
- Oikonomou, N., A. Thanasopoulou, A. Tzouveleakis, V. Harokopos, T. Paparountas, I. Nikitopoulou, W. Witke, A. Karameris, A. Kotanidou, D. Bouros, and V. Aidinis. 2009. Gelsolin expression is necessary for the development of modelled pulmonary inflammation and fibrosis. *Thorax* 64: 467–475.
- Bygdeman, M. 2003. Pharmacokinetics of prostaglandins. *Best Pract. Res. Clin. Obstet. Gynaecol.* 17: 707–716.
- Roulis, M., A. Kaklamanos, M. Scherthanner, P. Bielecki, J. Zhao, E. Kaffe, L.-S. Frommelt, R. Qu, M. S. Knapp, A. Henriques, et al. 2020. Paracrine orchestration of intestinal tumorigenesis by a mesenchymal niche. *Nature* 580: 524–529.
- Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth. 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43: e47.
- Carvalho, B. S., and R. A. Irizarry. 2010. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* 26: 2363–2367.
- Kauffmann, A., R. Gentleman, and W. Huber. 2009. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics* 25: 415–416.
- Stuart, T., A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, 3rd, Y. Hao, M. Stoeckius, P. Smibert, and R. Satija. 2019. Comprehensive integration of single-cell data. *Cell* 177: 1888–1902.e21.
- Reyffman, P. A., J. M. Walter, N. Joshi, K. R. Anekalla, A. C. McQuattie-Pimentel, S. Chiu, R. Fernandez, M. Akbarpour, C. I. Chen, Z. Ren, et al. 2019. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 199: 1517–1536.
- Yang, I. V., C. D. Coldren, S. M. Leach, M. A. Seibold, E. Murphy, J. Lin, R. Rosen, A. J. Neidermyer, D. F. McKean, S. D. Groshong, et al. 2013. Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis. *Thorax* 68: 1114–1121.
- Kim, S., J. D. Herazo-Maya, D. D. Kang, B. M. Juan-Guardela, J. Tedrow, F. J. Martinez, F. C. Scuirba, G. C. Tseng, and N. Kaminski. 2015. Integrative phenotyping framework (iPF): integrative clustering of multiple omics data identifies novel lung disease subphenotypes. *BMC Genomics* 16: 924.
- Oikonomou, N., M.-A. Mouratis, A. Tzouveleakis, E. Kaffe, C. Valavanis, G. Vilaras, A. Karameris, G. D. Prestwich, D. Bouros, and V. Aidinis. 2012. Pulmonary autotaxin expression contributes to the pathogenesis of pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* 47: 566–574.
- Bozyk, P. D., and B. B. Moore. 2011. Prostaglandin E2 and the pathogenesis of pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* 45: 445–452.
- Berthou, F., F. Ceppo, K. Dumas, F. Massa, B. Vergoni, S. Alemany, M. Cormont, and J. F. Tanti. 2015. The Tpl2 kinase regulates the COX-2/prostaglandin E2 axis in adipocytes in inflammatory conditions. *Mol. Endocrinol.* 29: 1025–1036.
- Eliopoulos, A. G., C. D. Dumitru, C. C. Wang, J. Cho, and P. N. Tschlis. 2002. Induction of COX-2 by LPS in macrophages is regulated by Tpl2-dependent CREB activation signals. *EMBO J.* 21: 4831–4840.
- Dackor, R. T., J. Cheng, J. W. Voltz, J. W. Card, C. D. Ferguson, R. C. Garrett, J. A. Bradbury, L. M. DeGraff, F. B. Lih, K. B. Tomer, et al. 2011. Prostaglandin

- E2 protects murine lungs from bleomycin-induced pulmonary fibrosis and lung dysfunction. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 301: L645–L655.
31. Failla, M., T. Genovese, E. Mazzon, M. Fruciano, E. Fagone, E. Gili, A. Barera, C. La Rosa, E. Conte, N. Crimi, et al. 2009. 16,16-Dimethyl prostaglandin E2 efficacy on prevention and protection from bleomycin-induced lung injury and fibrosis. *Am. J. Respir. Cell Mol. Biol.* 41: 50–58.
 32. Dumitru, C. D., J. D. Ceci, C. Tsatsanis, D. Kontoyiannis, K. Stamatakis, J. H. Lin, C. Patriotis, N. A. Jenkins, N. G. Copeland, G. Kollias, and P. N. Tschlis. 2000. TNF- α induction by LPS is regulated posttranscriptionally via a Tpl2/ERK-dependent pathway. *Cell* 103: 1071–1083.
 33. Rowley, S. M., T. Kuriakose, L. M. Dockery, T. Tran-Nguyen, A. D. Gingerich, L. Wei, and W. T. Watford. 2014. Tumor progression locus 2 (Tpl2) kinase promotes chemokine receptor expression and macrophage migration during acute inflammation. [Published erratum appears in 2020 *J. Biol. Chem.* 295: 7812.] *J. Biol. Chem.* 289: 15788–15797.
 34. Xiao, Y., J. Jin, M. Chang, M. Nakaya, H. Hu, Q. Zou, X. Zhou, G. C. Brittain, X. Cheng, and S. C. Sun. 2014. TPL2 mediates autoimmune inflammation through activation of the TAK1 axis of IL-17 signaling. *J. Exp. Med.* 211: 1689–1702.
 35. Vyrla, D., G. Nikolaidis, F. Oakley, M. J. Perugorria, P. N. Tschlis, D. A. Mann, and A. G. Eliopoulos. 2016. TPL2 kinase is a crucial signaling factor and mediator of NKT effector cytokine expression in immune-mediated liver injury. *J. Immunol.* 196: 4298–4310.
 36. Kannan, Y., J. Perez-Lloret, Y. Li, L. J. Entwistle, H. Khoury, S. Papoutsopoulou, R. Mahmood, N. R. Mansour, S. Ching-Cheng Huang, E. J. Pearce, et al. 2016. TPL-2 regulates macrophage lipid metabolism and M2 differentiation to control TH2-mediated immunopathology. *PLoS Pathog.* 12: e1005783.
 37. Serebrennikova, O. B., C. Tsatsanis, C. Mao, E. Gounaris, W. Ren, L. D. Siracusa, A. G. Eliopoulos, K. Khazaie, and P. N. Tschlis. 2012. Tpl2 ablation promotes intestinal inflammation and tumorigenesis in Apc^{min} mice by inhibiting IL-10 secretion and regulatory T-cell generation. *Proc. Natl. Acad. Sci. USA* 109: E1082–E1091.
 38. Wilborn, J., L. J. Crofford, M. D. Burdick, S. L. Kunkel, R. M. Strieter, and M. Peters-Golden. 1995. Cultured lung fibroblasts isolated from patients with idiopathic pulmonary fibrosis have a diminished capacity to synthesize prostaglandin E2 and to express cyclooxygenase-2. *J. Clin. Invest.* 95: 1861–1868.
 39. Coward, W. R., K. Watts, C. A. Feghali-Bostwick, A. Knox, and L. Pang. 2009. Defective histone acetylation is responsible for the diminished expression of cyclooxygenase 2 in idiopathic pulmonary fibrosis. *Mol. Cell. Biol.* 29: 4325–4339.
 40. Lama, V., B. B. Moore, P. Christensen, G. B. Toews, and M. Peters-Golden. 2002. Prostaglandin E2 synthesis and suppression of fibroblast proliferation by alveolar epithelial cells is cyclooxygenase-2-dependent. *Am. J. Respir. Cell Mol. Biol.* 27: 752–758.
 41. Maher, T. M., I. C. Evans, S. E. Bottoms, P. F. Mercer, A. J. Thorley, A. G. Nicholson, G. J. Laurent, T. D. Tetley, R. C. Chambers, and R. J. McAnulty. 2010. Diminished prostaglandin E2 contributes to the apoptosis paradox in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 182: 73–82.
 42. Mukherjee, S., W. Sheng, A. Michkov, K. Sriarm, R. Sun, A. Dvorkin-Gheva, P. A. Insel, and L. J. Janssen. 2019. Prostaglandin E₂ inhibits profibrotic function of human pulmonary fibroblasts by disrupting Ca²⁺ signaling. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 316: L810–L821.
 43. Moore, B. B., R. Paine, III, P. J. Christensen, T. A. Moore, S. Sitterding, R. Ngan, C. A. Wilke, W. A. Kuziel, and G. B. Toews. 2001. Protection from pulmonary fibrosis in the absence of CCR2 signaling. *J. Immunol.* 167: 4368–4377.
 44. Lavin, Y., D. Winter, R. Blecher-Gonen, E. David, H. Keren-Shaul, M. Merad, S. Jung, and I. Amit. 2014. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* 159: 1312–1326.
 45. Gibbings, S. L., R. Goyal, A. N. Desch, S. M. Leach, M. Prabagar, S. M. Atif, D. L. Bratton, W. Janssen, and C. V. Jakubzick. 2015. Transcriptome analysis highlights the conserved difference between embryonic and postnatal-derived alveolar macrophages. *Blood* 126: 1357–1366.
 46. Gautier, E. L., T. Shay, J. Miller, M. Greter, C. Jakubzick, S. Ivanov, J. Helft, A. Chow, K. G. Elpek, S. Gordonov, et al; Immunological Genome Consortium. 2012. Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat. Immunol.* 13: 1118–1128.
 47. Duffield, J. S., S. J. Forbes, C. M. Constandinou, S. Clay, M. Partolina, S. Vuthoori, S. Wu, R. Lang, and J. P. Iredale. 2005. Selective depletion of macrophages reveals distinct, opposing roles during liver injury and repair. *J. Clin. Invest.* 115: 56–65.
 48. Wynn, T. A., and K. M. Vannella. 2016. Macrophages in tissue repair, regeneration, and fibrosis. *Immunity* 44: 450–462.
 49. Desai, O., J. Winkler, M. Minasyan, and E. L. Herzog. 2018. The role of immune and inflammatory cells in idiopathic pulmonary fibrosis. *Front. Med. (Lausanne)* 5: 43.
 50. Williams, M., I. De Kleer, S. Henri, S. Post, L. Vanhoutte, S. De Prijck, K. Deswarte, B. Malissen, H. Hammad, and B. N. Lambrecht. 2013. Alveolar macrophages develop from fetal monocytes that differentiate into long-lived cells in the first week of life via GM-CSF. *J. Exp. Med.* 210: 1977–1992.
 51. Schneider, C., S. P. Nobs, M. Kurrer, H. Rehrauer, C. Thiele, and M. Kopf. 2014. Induction of the nuclear receptor PPAR- γ by the cytokine GM-CSF is critical for the differentiation of fetal monocytes into alveolar macrophages. *Nat. Immunol.* 15: 1026–1037.
 52. Misharin, A. V., L. Morales-Nebreda, G. M. Mutlu, G. R. Budinger, and H. Perlman. 2013. Flow cytometric analysis of macrophages and dendritic cell subsets in the mouse lung. *Am. J. Respir. Cell Mol. Biol.* 49: 503–510.
 53. Zaynagetdinov, R., T. P. Sherrill, P. L. Kendall, B. H. Segal, K. P. Weller, R. M. Tighe, and T. S. Blackwell. 2013. Identification of myeloid cell subsets in murine lungs using flow cytometry. *Am. J. Respir. Cell Mol. Biol.* 49: 180–189.
 54. Misharin, A. V., L. Morales-Nebreda, P. A. Reyfman, C. M. Cuda, J. M. Walter, A. C. McQuattie-Pimentel, C. I. Chen, K. R. Anekalla, N. Joshi, K. J. N. Williams, et al. 2017. Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. *J. Exp. Med.* 214: 2387–2404.
 55. Mould, K. J., L. Barthel, M. P. Mohning, S. M. Thomas, A. L. McCubbrey, T. Danhorn, S. M. Leach, T. E. Fingerlin, B. P. O'Connor, J. A. Reisz, et al. 2017. Cell origin dictates programming of resident versus recruited macrophages during acute lung injury. *Am. J. Respir. Cell Mol. Biol.* 57: 294–306.
 56. Landsman, L., and S. Jung. 2007. Lung macrophages serve as obligatory intermediate between blood monocytes and alveolar macrophages. *J. Immunol.* 179: 3488–3494.
 57. Gibbings, S. L., S. M. Thomas, S. M. Atif, A. L. McCubbrey, A. N. Desch, T. Danhorn, S. M. Leach, D. L. Bratton, P. M. Henson, W. J. Janssen, and C. V. Jakubzick. 2017. Three unique interstitial macrophages in the murine lung at steady state. *Am. J. Respir. Cell Mol. Biol.* 57: 66–76.
 58. Morris, A., K. Gibson, and R. G. Collman. 2014. The lung microbiome in idiopathic pulmonary fibrosis. What does it mean and what should we do about it? *Am. J. Respir. Crit. Care Med.* 190: 850–852.
 59. Ninou, I., E. Kaffé, S. Müller, D. C. Budd, C. S. Stevenson, C. Ullmer, and V. Aidinis. 2018. Pharmacologic targeting of the ATX/LPA axis attenuates bleomycin-induced pulmonary fibrosis. *Pulm. Pharmacol. Ther.* 52: 32–40.
 60. Fernandez, I. E., F. R. Greiffo, R. Roi, J. Behr, A. Forrest, and O. Eickelberg. 2018. Circulating MDSC modulate IPF progression by orchestrating immunosuppressive and pro-fibrotic networks. *Eur. Respir. J.* 52: OA4926.
 61. Liu, T., A. Rinke, K. Flaherty, and S. H. Phan. 2019. Potential role of myeloid-derived suppressor cells in pulmonary fibrosis. *J. Immunol.* 202: 182.3.
 62. Westphalen, K., G. A. Gusarova, M. N. Islam, M. Subramanian, T. S. Cohen, A. S. Prince, and J. Bhattacharya. 2014. Sessile alveolar macrophages communicate with alveolar epithelium to modulate immunity. *Nature* 506: 503–506.
 63. Biswas, S. K., and A. Mantovani. 2012. Orchestration of metabolism by macrophages. *Cell Metab.* 15: 432–437.
 64. Xie, N., Z. Tan, S. Banerjee, H. Cui, J. Ge, R. M. Liu, K. Bernard, V. J. Thannickal, and G. Liu. 2015. Glycolytic reprogramming in myofibroblast differentiation and lung fibrosis. *Am. J. Respir. Crit. Care Med.* 192: 1462–1474.
 65. Agostini, C., and C. Gurrieri. 2006. Chemokine/cytokine cocktail in idiopathic pulmonary fibrosis. *Proc. Am. Thorac. Soc.* 3: 357–363.
 66. Van Acker, G. J., G. Perides, E. R. Weiss, S. Das, P. N. Tschlis, and M. L. Steer. 2007. Tumor progression locus-2 is a critical regulator of pancreatic and lung inflammation during acute pancreatitis. *J. Biol. Chem.* 282: 22140–22149.

Increased Autotaxin Levels in Severe COVID-19, Correlating with IL-6 Levels, Endothelial Dysfunction Biomarkers, and Impaired Functions of Dendritic Cells

Αυξημένα επίπεδα αυτοταξίνης σε σοβαρή νόσο COVID-19, συσχετιζόμενα με επίπεδα IL-6, βιοδείκτες ενδοθηλιακής δυσλειτουργίας και προβληματικές λειτουργίες δενδριτικών κυττάρων

Η αυτοταξίνη (ATX; *ENPP2*) είναι μία εκκρινόμενη λισοφωσφολιπάση D που καταλύει την εξωκυττάρια παραγωγή λισοφωσφατιδικού οξέος (LPA). Προηγούμενες γενετικές και φαρμακολογικές μελέτες έχουν καταδείξει τον παθολογικό ρόλο της ATX και της σηματοδότησης του LPA στον τραυματισμό, την φλεγμονή και την ίνωση των πνευμόνων. Στον παρόν άρθρο, αυξημένα επίπεδα μεταφραστικού RNA του *ENPP2* ανιχνεύθηκαν σε ανοσοποιητικά κύτταρα ρινοφαρυγγικών επιχρισμάτων ασθενών με COVID-19. Αυξημένα επίπεδα ATX ορού βρέθηκαν σε σοβαρά περιστατικά COVID-19. Τα επίπεδα ATX ορού συσχετίστηκαν με τα αντίστοιχα επίπεδα ορού της IL-6, καθώς και με βιοδείκτες ενδοθηλιακής καταστροφής, προτείνοντας την αλληλεπίδραση του άξονα ATX/LPA με περιστατικά υπερφλεγμονώδους συνδρόμου και συσχετιζόμενης αγγειακής δυσλειτουργίας στην COVID-19. Αντιστοίχως, σύμφωνα με δεδομένα από δύο ανεξάρτητες κοόρτες, χορήγηση δεξαμεθαζόνης (Dex) σε ασθενείς υπο μηχανική οξυγόνωση μείωσε τα επίπεδα της ATX, υποδεικνύοντας πως τα θεραπευτικά οφέλη της Dex συμπεριλαμβάνουν την καταστολή της ATX. Επιπροσθέτως, μεγάλης κλίμακας ανάλυση πολλαπλών συνόλων δεδομένων αλληλούχισης RNA από μεμονωμένα κύτταρα αποκάλυψε το μεταγραφωμικό τοπίο του *ENPP2* στην COVID-19. Επίσης, πρότεινε την συμμετοχή της ATX στην ομοίωση των δενδριτικών κυττάρων, τα οποία επιδεικνύουν τόσο αριθμητικά όσο και λειτουργικά ελλείματα κατά την COVID-19. Ως εκ τούτου, η ATX πιθανώς να έχει λειτουργικό χαρακτήρα στην παθογένεση της COVID-19, προτείνοντας τη φαρμακολογική στόχευση του εν λόγω μορίου ως μια επιπρόσθετη πιθανή θεραπευτική επιλογή, τόσο κατά τη διάρκεια όσο και μετά την ενδονοσοκομειακή νοσηλεία των ασθενών.



Article

Increased Autotaxin Levels in Severe COVID-19, Correlating with IL-6 Levels, Endothelial Dysfunction Biomarkers, and Impaired Functions of Dendritic Cells

Ioanna Nikitopoulou ^{1,†}, Dionysios Fanidis ^{2,†} , Konstantinos Ntatsoulis ^{2,†} , Panagiotis Moulos ³ , George Mpekoulis ⁴, Maria Evangelidou ⁵, Alice G. Vassiliou ¹ , Vasiliki Dimakopoulou ⁶, Edison Jahaj ⁷ , Stamatiou Tsipilis ⁷ , Stylianos E. Orfanos ⁷, Ioanna Dimopoulou ⁷, Emmanouil Angelakis ⁵, Karolina Akinosoglou ⁶, Niki Vassilaki ⁴, Argyrios Tzouvelekis ⁶, Anastasia Kotanidou ^{1,7} and Vassilis Aidinis ^{2,*}



Citation: Nikitopoulou, I.; Fanidis, D.; Ntatsoulis, K.; Moulos, P.; Mpekoulis, G.; Evangelidou, M.; Vassiliou, A.G.; Dimakopoulou, V.; Jahaj, E.; Tsipilis, S.; et al. Increased Autotaxin Levels in Severe COVID-19, Correlating with IL-6 Levels, Endothelial Dysfunction Biomarkers, and Impaired Functions of Dendritic Cells. *Int. J. Mol. Sci.* **2021**, *22*, 10006. <https://doi.org/10.3390/ijms221810006>

Academic Editor: Anindita Das

Received: 25 August 2021

Accepted: 11 September 2021

Published: 16 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

- ¹ GP Livanos and M Simou Laboratories, 1st Department of Critical Care & Pulmonary Services, Medical School, National & Kapodistrian University of Athens, Evangelismos General Hospital, 10676 Athens, Greece; joannaniki@gmail.com (I.N.); alvass75@gmail.com (A.G.V.); akotanid@med.uoa.gr (A.K.)
 - ² Institute of Bio-Innovation, Biomedical Sciences Research Center Alexander Fleming, 16672 Athens, Greece; fanidis@fleming.gr (D.F.); ntatsoulis@fleming.gr (K.N.)
 - ³ Institute for Fundamental Biomedical Research, Biomedical Sciences Research Center Alexander Fleming, 16672 Athens, Greece; moulos@fleming.gr
 - ⁴ Molecular Virology Laboratory, Department of Diagnostics, Hellenic Pasteur Institute, 11521 Athens, Greece; g.mpekoulis@pasteur.gr (G.M.); nikiv@pasteur.gr (N.V.)
 - ⁵ Department of Diagnostics, Hellenic Pasteur Institute, 11521 Athens, Greece; meuangelidou@pasteur.gr (M.E.); e.angelakis@pasteur.gr (E.A.)
 - ⁶ Department of Respiratory Medicine, University Hospital of Patras, 26504 Patras, Greece; dimakopoulou.vasilina@gmail.com (V.D.); akin@upatras.gr (K.A.); atzouvelekis@upatras.gr (A.T.)
 - ⁷ 1st Department of Critical Care & Pulmonary Services, Medical School, National & Kapodistrian University of Athens, Evangelismos General Hospital, 10676 Athens, Greece; edison.jahaj@gmail.com (E.J.); stamostsipil@gmail.com (S.T.); stylianosorfanosuoa@gmail.com (S.E.O.); idimo@otenet.gr (I.D.)
- * Correspondence: V.Aidinis@Fleming.gr; Tel.: +302-109-654-382
† Equal contribution.

Abstract: Autotaxin (ATX; *ENPP2*) is a secreted lysophospholipase D catalyzing the extracellular production of lysophosphatidic acid (LPA), a pleiotropic signaling phospholipid. Genetic and pharmacologic studies have previously established a pathologic role for ATX and LPA signaling in pulmonary injury, inflammation, and fibrosis. Here, increased *ENPP2* mRNA levels were detected in immune cells from nasopharyngeal swab samples of COVID-19 patients, and increased ATX serum levels were found in severe COVID-19 patients. ATX serum levels correlated with the corresponding increased serum levels of IL-6 and endothelial damage biomarkers, suggesting an interplay of the ATX/LPA axis with hyperinflammation and the associated vascular dysfunction in COVID-19. Accordingly, dexamethasone (Dex) treatment of mechanically ventilated patients reduced ATX levels, as shown in two independent cohorts, indicating that the therapeutic benefits of Dex include the suppression of ATX. Moreover, large scale analysis of multiple single cell RNA sequencing datasets revealed the expression landscape of *ENPP2* in COVID-19 and further suggested a role for ATX in the homeostasis of dendritic cells, which exhibit both numerical and functional deficits in COVID-19. Therefore, ATX has likely a multifunctional role in COVID-19 pathogenesis, suggesting that its pharmacological targeting might represent an additional therapeutic option, both during and after hospitalization.

Keywords: COVID-19; ARDS; cytokine storm; vascular dysfunction; pulmonary fibrosis; autotaxin (ATX; *ENPP2*); lysophosphatidic acid (LPA); dendritic cells (DCs)

1. Introduction

The leading symptom of COVID-19, beyond cough and fever, is hypoxemia, leading to dyspnea in severe cases, attributed to impaired lung mechanics and/or vasoconstriction [1,2]. Endothelial dysfunction is a major characteristic of COVID-19 [3], shared with hypertension, diabetes, and obesity, the most common comorbidities that are associated with poor prognosis [1,2]. The respiratory epithelial cell damage that follows viral infection and replication stimulate, depending on the underlying genetic and metabolic context, a hyperinflammatory state denominated “cytokine storm” [4]. The excessive production of pro-inflammatory cytokines, such as TNF and IL-6, further induces endothelial damage and lung injury, and its more severe form, Acute Respiratory Distress Syndrome (ARDS), that can result in respiratory and/or multi-organ failure and death [5].

A subset of surviving COVID-19 ARDS-like patients will develop a fibroproliferative response characterized by fibroblast accumulation and ECM deposition [6], also evident in postmortem histopathological analysis of the lungs of COVID-19 patients [7]. Moreover, many discharged COVID-19 patients present with abnormally pulmonary architecture and functions [8–12], suggesting persisting fibrotic abnormalities, pending long-term follow up studies. Single-cell RNA sequencing (scRNAseq) analysis and transcriptional profiling indicated similarities in expression profiles between idiopathic pulmonary fibrosis (IPF) and COVID-19 [13,14], while CoV-2 infection has been suggested to stimulate the expression of major pro-fibrotic factors including TGF β [15]. Vice versa, patients with interstitial lung diseases (ILD) had an increased risk for severe COVID-19 and poor outcomes (ICU admittance, death) following CoV-2 infection [16–18].

Autotaxin (ATX; ENPP2) is a secreted lysophospholipase D that can be found in most biological fluids, including blood and bronchoalveolar lavage fluid (BALF), largely responsible for the extracellular production of lysophosphatidic acid (LPA), a growth factor-like signaling phospholipid. Increased ATX expression and LPA signaling has been reported in cancer as well as in chronic inflammatory diseases [19], including IPF [20,21]. Genetic and pharmacologic studies have further uncovered a therapeutic potential for ATX in IPF [20,22–24], leading to phase III clinical trials [25]. Given the associations of COVID-19 with pulmonary fibrosis, the pro-fibrotic properties of ATX, as well the many reported LPA effects on pulmonary cells and especially the vasculature [26], in this study we explored a possible association of ATX with COVID-19. In this context, we quantified *ENPP2* mRNA levels in nasopharyngeal swabs and ATX protein levels in the sera of two cohorts of COVID-19 patients, while we performed a large-scale analysis of recently published scRNAseq COVID-19 datasets.

2. Results

2.1. Increased *ENPP2* mRNA Levels in Nasopharyngeal Swab Samples from COVID-19 Patients

As viral infections have been reported to stimulate *ENPP2* mRNA expression [27,28], and to examine if CoV-2 infection has similar effects, we first quantified *ENPP2* mRNA levels with Q-RT-PCR in nasopharyngeal swab samples (Table 1). A significant increase was found in *ENPP2* mRNA expression in mild and severe COVID-19 patients, as compared to non-infected subjects (Figure 1). Therefore, CoV-2 infection stimulates *ENPP2* mRNA expression in the respiratory epithelial or immune cells that compose the nasopharyngeal swab samples.

Table 1. Increased *ENPP2* mRNA expression in nasopharyngeal swabs from COVID-19 patients compared to healthy, non-infected controls.

	Negative Healthy	Positive Mild	Positive Severe/Critical
Number of patients (n)	21	21	21
ATX ($2^{-\Delta\Delta C_t}$, mean \pm SD)	2.15 \pm 1.37	5.38 \pm 2.34 ****	5.76 \pm 2.18 ****
Sex			
Male	8 (38.09%)	10 (47.6%)	17 (80.95%)
Female	13 (61.9%)	9 (42.8%)	4 (19.04%)
Not recorded	0	2 (9.52%)	0
Age (years, mean \pm SD)	50.15 \pm 20.86	37.78 \pm 11.89	63.38 \pm 17.23

Statistical significance was assessed with one-way ANOVA followed by Bonferroni post hoc correction; **** denotes $p < 0.0001$. ATX values are presented at Figure 1.

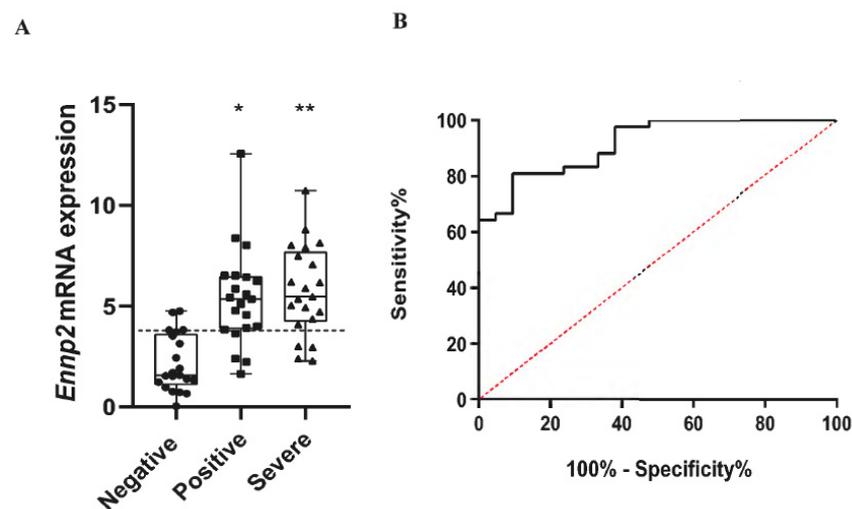


Figure 1. Increased *ENPP2* mRNA expression in nasopharyngeal swab samples from patients with mild or severe/critical COVID-19. (A) *ENPP2* mRNA values (mean $2^{-\Delta\Delta C_t}$) from the two groups of patients ($n = 21$) and the control group ($n = 21$). The horizontal dotted line indicates the optimal threshold value (cut-off). Data are represented as box plots; line in the middle, median; box edges, 25th to 75th centiles; whiskers, range of values. p values were calculated with the non-parametric Mann–Whitney U test. (B) ROC curves were generated after merging the results for the two positive groups (mild and severe/critical), and AUC, 95% CI, p values, and cut-off points with their specificity and sensitivity were calculated. The dotted line (in red) represents perfect chance (positive likelihood ratio=sensitivity/(1-specificity)=1). Positive mild versus negative samples: * $p = 4 \times 10^{-6}$, positive critical/severe versus negative samples: ** $p = 2.92 \times 10^{-7}$.

2.2. Increased Serum ATX Protein Levels in Severe COVID-19 Patients

To examine if systemic levels of ATX are possibly increased upon COVID-19, ATX was quantified with an ELISA kit in the serum of COVID-19 patients hospitalized at the Evangelismos University Hospital (Table 2; completely independent from cohort/Table 1). The cohort consisted of both WARD ($n = 47$; no Dex treatment), as well as of Intensive Care Unit (ICU) patients, which were further separated in patients receiving dexamethasone (Dex) treatment ($n = 37$) or not (NO Dex; $n = 32$). A large proportion of patients suffered from comorbidities and were receiving a variety of medications prior to admission, while COVID-19-targeted treatment included azithromycin, chloroquine, and lopinavir/ritonavir in different combinations per WHO recommendations at that time (Table 2). In comparison with WARD patients, ICU patients were hypoxemic (low ratio of arterial oxygen partial pressure to fractional inspired oxygen; PaO_2/FiO_2), lymphopenic (low lymphocyte numbers), and had increased LDH levels (Table 2), all three suggested as disease severity markers.

Table 2. Clinical characteristics and laboratory data of COVID-19 patients hospitalized at the Evangelismos general hospital.

	WARD No Dex	ICU No Dex	ICU + Dex
Number of patients (n)	47	37	32
ATX (ng/mL, mean ± SD)	310.32 ± 98.85 *	443 ± 172.90	246.15 ± 73.74 *
Sex			
Male	33 (70.21%)	31 (83.78%)	22 (62.5%)
Female	14 (29.78%)	6 (16.21%)	10 (31.25%)
Age (years, mean ± SD)	54.63 ± 15.46	63.54 ± 10.89	65.5 ± 10.7
Comorbidities n (%)			
Hypertension	13 (27.65%)	17 (45.94%)	12 (37.5%)
Diabetes	4 (8.51%)	5 (13.51%)	5 (15.62%)
Coronary artery disease	8 (17.02%)	4 (10.81%)	4 (12.5%)
COPD	1 (2.12%)	1 (2.7%)	2 (6.25%)
Asthma	2 (4.25%)	1 (2.7%)	1 (3.12%)
Hyperlipidemia	9 (19.14%)	9 (24.32%)	8 (25%)
Hepatitis	0 (0%)	1 (2.7%)	0 (0%)
COVID-19 treatment	0	11	
Azithromycin/chloroquine/lopinavir/ritonavir	6	7	
Azithromycin/chloroquine	0	2	
Lopinavir/ritonavir/chloroquine	0	3	
Chloroquine Plasma	0	1	
Clinical measurements			
Mean arterial pressure (mmHg)	83.19 ± 8.86	82.83 ± 16.52	77.55 ± 8.54
PaO ₂ /FiO ₂ (mmHg)	301.5 ± 79.81 *	194.86 ± 86.64	85.94 ± 15.97 *
Glucose (mg/dL)	133.5 ± 113.3	164.53 ± 77.73	164.06 ± 75.40
Creatinine (mg/dL)	0.9 ± 0.33	1.02 ± 0.32	0.95 ± 0.72
CRP (mg/dL)	6.8 ± 8.96	14 ± 10.17	13.83 ± 9.6
Total bilirubin (mg/dL)	0.5 ± 0.33	0.73 ± 0.5	0.61 ± 0.29
White blood cell count (per µL)	6995 ± 3468	10,125 ± 4633	11,705 ± 10,372
Neutrophils (%)	69.34 ± 13.51	81.34 ± 6.64	83.12 ± 12.2
Lymphocytes (%)	24.03 ± 10.89 *	12.63 ± 5.63	11.12 ± 11.23
Platelets (per µL)	240,297 ± 110,028	237,783 ± 101,338	257,000 ± 79,581
INR (median IQR)	1.06 ± 0.09	2.07 ± 5.73	1.26 ± 0.65
D-dimer (pg/mL)	1.19 ± 1.72	0.47 ± 0.26	1.39 ± 0.93
AST (IU/L)	36.65 ± 30.65	54.18 ± 39.95	121.4 ± 329.9
ALT (IU/L)	33.15 ± 23.58	45.9 ± 28.08	60.8 ± 72.4
LDH (U/L)	286.36 ± 122.08 *	498.48 ± 242.34	591.23 ± 490.84
Fibrinogen (mg/dL)	514.06 ± 176.18	638.18 ± 158.76	630.3 ± 172.2
Ferritin (pg/mL)	513.48 ± 815.55	2786 ± 694.48	912.47 ± 826.91
APACHE II score	5.25 ± 2.94	14.27 ± 5.08	15.4 ± 3.89
SOFA score	2 ± 1	6.83 ± 3.08	5.4 ± 1.81

Statistical significance with ICU NO Dex group values was assessed with one-way ANOVA followed by Bonferroni post hoc correction; * denotes $p < 0.0001$. ATX values appear at Figures 2, 3, S1 and S2.

Increased ATX serum concentrations were discovered in ICU patients (not receiving Dex) as compared with WARD patients (Figure 2A), suggesting a possible association of ATX with disease severity. However, no substantial, statistically significant correlation was observed independently with the applicable severity markers (data not shown and Table 2); no statistically significant differences of ATX levels between the sex or the comorbidities of COVID-19 patients was detected (Figure S1).

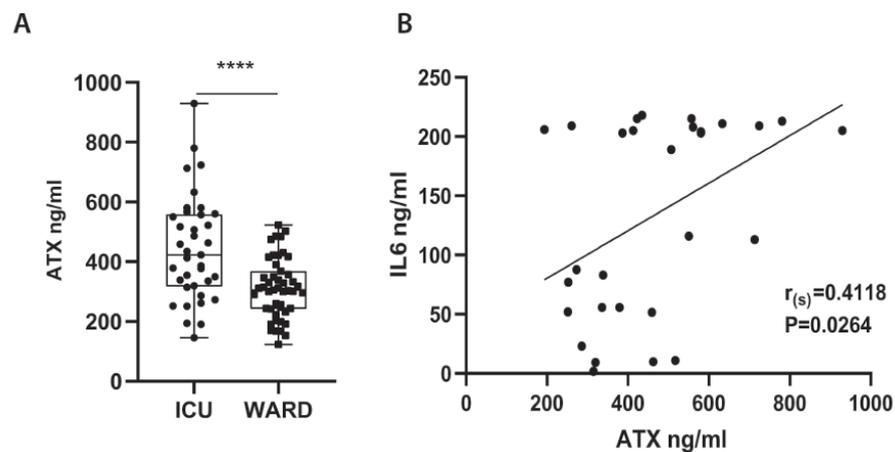


Figure 2. Increased serum ATX protein levels in COVID-19 patients hospitalized in the intensive care unit (ICU), correlating with increased IL-6 levels. (A) ATX protein levels were measured with a commercial ELISA kit in the sera of COVID-19 patients hospitalized (without Dex treatment) in the COVID-19 Ward ($n = 47$) or the ICU ($n = 37$) of the Evangelismos hospital. Statistical significance, given the normal distribution of values, was assessed with an unpaired *t*-test. **** denotes $p < 0.0001$. (B) ATX serum levels correlated with serum IL-6 levels ($n = 29$), as assessed with Spearman correlation ($r_{(s)}$).

Among the different cytokines that have been reported, with great variance, to get elevated in the COVID-19-induced cytokine storm, IL-6 was found to be the most predictive one [29,30], while in this cohort only IL-6 could be detected in high amounts; the IL-6 levels in some patients reached the upper detection threshold (Figure 2B). Moreover, IL-6 has been reported to stimulate ATX expression in different contexts [31,32]. Therefore, we next examined if ATX and IL-6 serum levels correlate, to discover if, most importantly, ATX levels correlated significantly with IL-6 levels in the serum of ICU patients (not receiving Dex) (Figure 2B), suggesting a possible interplay of ATX/LPA with the cytokine storm in COVID-19.

ICU non-survivors in this cohort had higher levels of the endothelial dysfunction markers soluble E-selectin (sE-sel), soluble P-selectin (sP-sel), soluble intercellular adhesion molecule 1 (sICAM-1), and angiopoietin 2 (ANG-2) when compared to survivors, as recently reported using a subset of the current Evangelismos cohort samples [33]. Interestingly, the increased ATX protein levels correlated with the increased protein levels of sE-sel and sICAM (Figure S2) in ICU patients, suggesting a role for ATX/LPA in COVID-19-induced endothelial dysfunction.

2.3. Dex Therapeutic Effects in COVID-19 Include the Suppression of ATX Serum Levels

The first line of therapy for many inflammatory diseases as well as respiratory infections is Dex, which lowers the expression of pro-inflammatory cytokines including IL-6, and which has been proven effective in COVID-19 patients requiring, invasive or not, oxygenation [34,35]. Therefore, we next examined ATX serum levels in intubated, or not, ICU patients receiving, or not, Dex treatment. Remarkably, Dex treatment was discovered to potently suppress ATX serum levels in ventilated patients (Figure 3A), while intubated ICU patients receiving no Dex presented with the highest overall ATX serum levels. Identical results were obtained in another cohort of ICU patients from the University hospital of Patras (Table 3) (Figure 3B), indicating that the therapeutic benefits of Dex include the suppression of ATX serum levels.

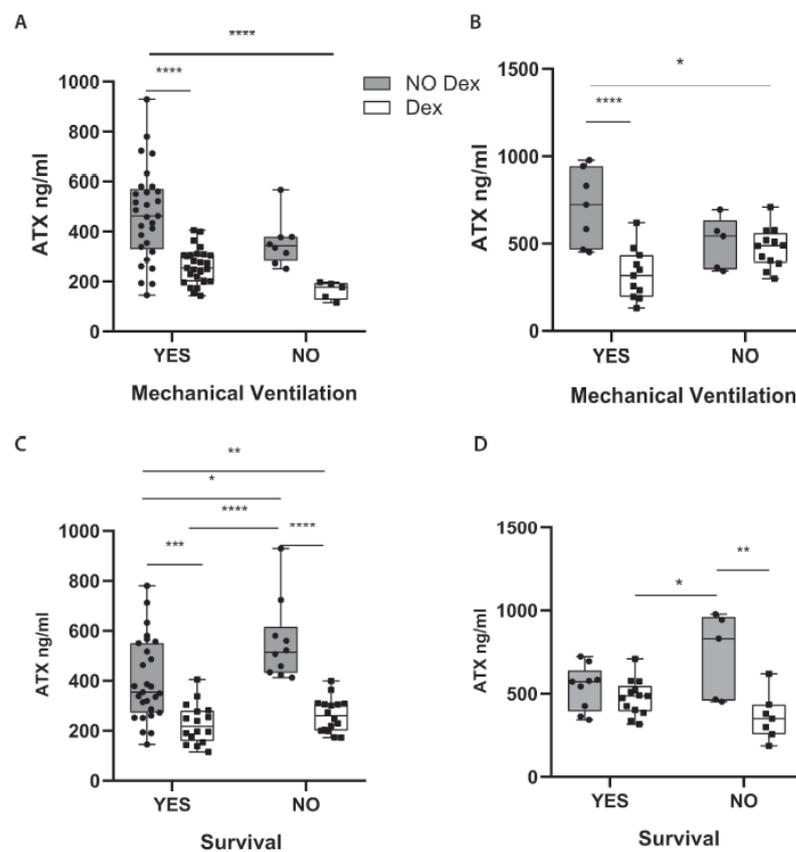


Figure 3. Dexamethasone therapeutic effects include the suppression of ATX serum levels. ATX protein levels were measured, with a commercial ELISA kit in the serum of COVID-19 patients hospitalized in the ICU of (A,C) the Evangelismos or (B,D) the Patras hospital. The measurements in the Dex groups in 3A and C, are the same as in Figure 2A. Statistical significance, given the normal distribution of values, was assessed with 2-way ANOVA followed by Bonferroni post hoc correction. *, **, ***, **** denote $p < 0.05$, $p < 0.01$, $p < 0.001$ and $p < 0.0001$, respectively.

Table 3. Clinical characteristics and laboratory data of COVID-19 patients hospitalized at the ICU of the University Hospital of Patras.

	ICU No Dex	ICU + Dex
Number of patients (n)	12	23
ATX (ng/mL, mean \pm SD)	624.36 \pm 203.5	404.16 \pm 145.5 **
Sex		
Male	9 (75%)	18 (78.26%)
Female	3 (25%)	5 (21.73%)
Age (years, mean \pm SD)	66.75 \pm 13.31	59.43 \pm 15.42
Comorbidities n (%)		
Hypertension	5 (41.6%)	10 (43.47%)
Diabetes	0 (0%)	4 (17.39%)
Coronary artery disease	2 (16.6%)	1 (4.34%)
COPD	0 (0%)	2 (8.69%)
Asthma	0 (0%)	0 (0%)
Hyperlipidemia	3 (25%)	6 (26.08%)
Hepatitis	(%)	(%)

Table 3. Cont.

	ICU No Dex	ICU + Dex
COVID 19 treatment		
Azithromycin/chloroquine/lopinavir/ritonavir	4 (33.33%)	0 (0%)
Azithromycin/chloroquine	6 (50%)	5 (21.73%)
Lopinavir/ritonavir/chloroquine	1 (8.33%)	0 (0%)
Chloroquine	0 (%)	0 (%)
Plasma	0 (%)	0 (%)
Clinical measurements		
Glucose (mg/dL)	120.5 ± 24.57	154.04 ± 47.98 *
Creatinine (mg/dL)	0.95 ± 0.42	0.93 ± 0.38
CRP (mg/dL)	8.53 ± 5.51	15.55 ± 12.68
Total bilirubin (mg/dL)	0.67 ± 0.27	0.7 ± 0.36
Lymphocytes (absolute number)	0.62 ± 0.35	0.83 ± 0.5
INR (median IQR)	1.09 ± 0.13	1.07 ± 0.12
D-dimer (pg/mL)	2.21 ± 2.14	1.45 ± 1.72
LDH (U/L)	370 ± 129	455.08 ± 188.56
Fibrinogen (mg/dL)	630.55 ± 168.1	545 ± 182.32
Ferritin (pg/mL)	950 ± 382.64	1131.09 ± 1223.55

Statistical significance was assessed with an unpaired *t*-test; *, ** denotes $p < 0.05$, $p < 0.01$. ATX values appear in Figure 3.

Moreover, ATX levels in ICU patients not receiving Dex treatment negatively affected survival, and non-surviving ICU patients receiving no Dex presented with the higher overall ATX serum levels (Figure 3C,D).

2.4. The ENPP2 Expression Landscape in COVID-19

To identify possible ATX expressing cells in the nasopharyngeal swab (NS) samples (Figure 1), peripheral blood monocytes (PBMCs) in the circulation (Figures 2 and 3), as well as in BALF and lung tissue cells, we re-analyzed and mined several scRNAseq datasets of COVID-19 patients and healthy controls, from recent high impact studies (Table 1), collectively interrogating the gene expression of more than 10^6 cells; cell clustering and naming followed that of the original analyses, which both varied between studies/datasets.

In NS cells, ATX was found to be mainly expressed by natural killer cells (NKs) and monocyte-derived macrophages (MoAM) (Figures 4A and S3A), as detected in two NS datasets (Table 1). In the circulation, and in both PBMCs datasets (Table 1), *ENPP2* expression was mainly detected, remarkably, in plasmacytoid DCs (pDCs; Figures 4B and S3B). In BALF cells (Table 1), *ENPP2* expression was also mainly detected in pDCs, as well as MoAMs (Figure 4C and Figure S3C). In lung tissue (Table 1), *ENPP2* was found to be primarily expressed in arterial and mesothelial cells, as well as in cells of the monocytic lineage (Figure 4D and Figure S3D). A similar lung tissue profile was also detected (Figure S3E) in an IPF scRNAseq dataset (Table 1), extending the similarities of pathogenic mechanisms between IPF and COVID-19 and supporting a common role for ATX.

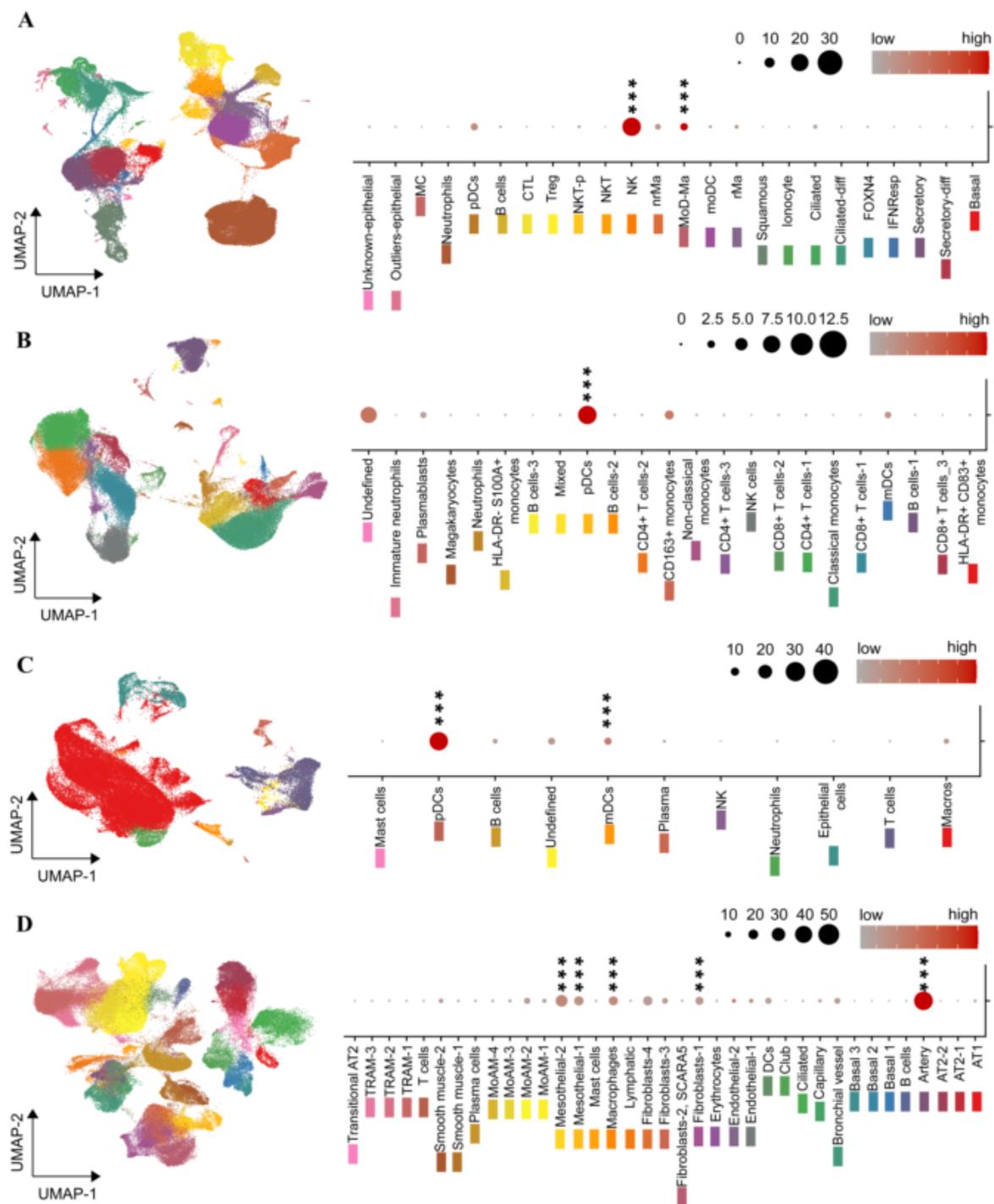


Figure 4. The *ENPP2* expression landscape in COVID-19. *ENPP2* expression was assessed in four datasets of COVID-19/healthy control datasets, each representing a different sampling site ((A) nasopharyngeal swabs; (B) PBMC; (C) BALF; (D) lung tissue; Table 1). UMAP plots (on the left) depict the cellular composition of these sites, while dot plots (on the right) the expression pattern of *ENPP2* in the detected cell types. Dot plots cell type color coding refers to that of the respective UMAP. Marker genes, denoted by stars, were detected using a Wilcoxon rank sum test; FC > 1.2, Bonferroni corrected $p < 0.05$; *** denotes $p < 0.01$ (PMIDs: (A) 32591762; (B) 32810438; (C) 32398875; (D) 33257409; Table S1).

2.5. A Role for ATX in the Homeostasis of Dendritic Cells?

Given the *ENPP2* expression from monocytic cells and especially pDCs, we next interrogated *ENPP2* mRNA levels specifically in pDCs from COVID-19 patients in comparison with control samples, subsets of the datasets analyzed in Figure 4. Confined by the

limited numbers of lung pDCs, as well as the detected genes per cell and the relatively low expression levels of *ENPP2*, the analysis indicated a statistically significant overexpression of *ENPP2* in COVID-19 circulating pDCs (Figure 5B). Noteworthy, DCs are the highest *ENPP2* expressing immune cells during healthy conditions, as identified upon querying a large-scale RNAseq dataset interrogating gene expression of 28 immune cell types (79 healthy volunteers and 337 patients from 10 immune-related diseases) [36] (Figure S4A). Similar analysis indicated that the main LPA receptor expressed by DCs, among the at least six LPA receptors reported thus far [37], is LPAR2 (Figure S4B), which has been suggested to convey anti-inflammatory LPA signals to DCs [38]. Furthermore, increased *ENPP2* mRNA expression was detected in pDCs from patients with systemic lupus erythematosus (SLE), adult-onset Still's disease (AOSD), mixed connective tissue disease (MCTD), and idiopathic inflammatory myopathy (IIM) than in DCs from healthy volunteers (Figure S4C), suggesting that overexpression of *ENPP2* in pDCs may be a common theme in inflammation.

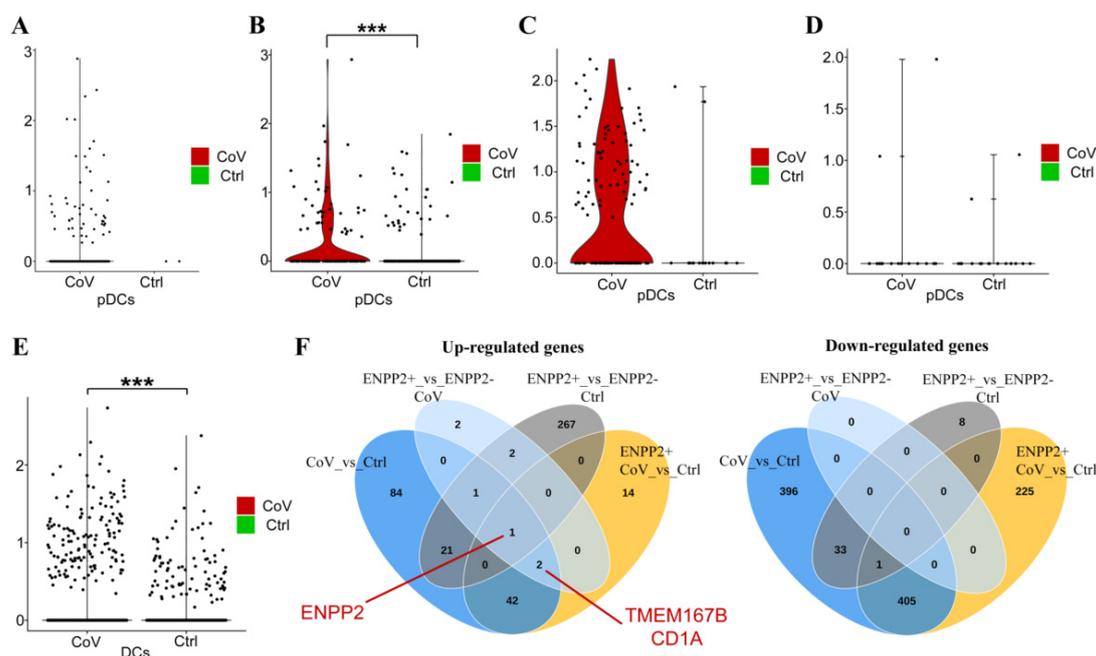


Figure 5. *ENPP2* mRNA expression is upregulated in peripheral pDCs and lung tissue DCs of COVID-19 patients. (A–D) Differential expression of *ENPP2* in pDCs of COVID-19 patients versus healthy controls: (A) nasopharyngeal swabs; (B) PBMC; (C) BALF; (D) lung tissue. *ENPP2* mRNA expression was found up-regulated in peripheral (B) pDCs of COVID-19 patients. (E) Differential expression analysis indicates increased *ENPP2* mRNA expression in COVID-19 lung tissue DCs compared to healthy controls. (F) Venn diagrams of deregulated genes in lung DCs. Differential expression was performed using a Wilcoxon rank sum test; FC > 1.2 and Bonferroni corrected $p < 0.05$; *** denotes Bonferroni adjusted $p < 0.01$. (PMIDs: (A) 32591762; (B) 32810438; (C) 32398875; (D) 33257409; Table S1).

Finally, and to gain mechanistic insights into the possible role of ATX in DC homeostasis upon COVID-19, we first analyzed differential gene expression in COVID-19 DCs (as pDCs were too few), from the only COVID-19 lung dataset [13] allowing such analysis, as well as in *ENPP2*-expressing (*ENPP2*⁺) DCs (Table S5). Increased *ENPP2* expression was also detected in all lung DCs (Figure 5E), while comparative analysis (Venn diagrams Figure 5F) highlighted two genes upregulated in *ENPP2*⁺ COVID-19 DCs, transmembrane protein 176B (TMEM176B) and CD1a, that have been both proposed as DC differentiation and/or maturation markers [39–43], suggesting that *ENPP2* expression may modulate DC homeostasis.

3. Discussion

Previous studies have shown that HCV, HIV, and HBV viruses increase *Enpp2* mRNA expression in infected cells and/or raise systemic ATX levels [27,44,45]. As shown here, increased *ENPP2* mRNA expression was detected in nasopharyngeal swab samples from COVID-19 patients in comparison to non-infected healthy controls (Figure 1), while scRNAseq re-analysis revealed that the highest *ENPP2* expressing cells in swabs are immune cells (Figures 5A and S5A), suggesting that CoV-2 infection stimulates *ENPP2* expression from immune cells in the nasopharynx. LPA, the enzymatic product of ATX and its effector molecule, has been shown to directly affect HCV viral infection and replication [27,28], suggesting that a similar autocrine mode of action may be in play in COVID-19, where ATX produced by the infected host cell would stimulate local LPA production, in turn facilitating viral entry and replication.

Increased serum ATX protein have been reported in cancer, liver diseases, as well as respiratory diseases including asthma and pulmonary fibrosis [19,46], while increased levels of serum ATX were recently reported in ARDS [47]. Here, increased ATX sera levels were detected in ICU-hospitalized COVID-19 patients (receiving no Dex treatment) compared to patients with less severe disease (Figure 2), suggesting increased ATX expression as an additional commonality of ARDS and COVID-19.

Several studies have indicated a deregulated serum lipid profile of COVID-19 patients, mainly focusing on fatty acids and triglycerides (reviewed in [48]); however, little is known on phospholipid homeostasis upon CoV-2 infection. As ATX is largely responsible for extracellular LPA synthesis, it would be complementary to quantify LPA levels in the same samples of this study. However, the analyzed serum samples cannot be used because blot clotting and platelet activation stimulate massive LPA release [49]. Moreover, all samples should have been collected in siliconized tubes, to avoid the known attachment of lipids to tubing, and kept at -80°C or lower without repeated freeze thawing [48]. Therefore, a new, multi-controlled, perspective study will be necessary to assess the levels of the different LPA species, as well as related phospholipids, such as LPC, LPE, LPS, and S1P, with whom LPA shares common interconnected biosynthetic pathways.

The origin of serum ATX is not completely deciphered; however, >40% of mouse serum constitutively active ATX has been suggested to originate from the adipose tissue [50], which was shown to be able to modulate the pathophysiology of distant metabolically active organs [51,52]. Moreover, serum ATX has been reported to correlate with insulin resistance in older humans with obesity [53], while mice with heterozygous *Enpp2* deficiency were protected from HFD-induced obesity and systemic insulin resistance [52]. Several additional reports have incriminated the ATX/LPA axis in the regulation of glucose homeostasis and insulin resistance (reviewed in [54]), among the main comorbidities of COVID-19, suggesting adipose tissue-derived ATX as a possible pathologic link between obesity and COVID-19. No correlation of ATX serum levels with the related underlying comorbidities of COVID-19 patients, cardiovascular diseases, diabetes, and dyslipidemia (Figure S1B) or the body mass index (BMI; data not shown) of patients was observed. However, the lack of correlation could be due to the heterogeneity and size of the examined cohorts, as well as due to the corresponding treatments the patients were receiving for their underlying pathologic conditions prior to CoV-2 infection and hospitalization. In particular, several reports have associated dyslipidemia with increased severity and mortality of COVID-19 (reviewed in [55]), which led to suggestions for lipid-lowering therapies of COVID-19 patients [56], including the administration of statins [57], that inhibit HMG-CoA reductase—a rate-limiting enzyme for cholesterol synthesis, to COVID-19 patients. However, results from clinical trials have been controversial [58–60], and larger studies are needed to reach safe conclusions. Interestingly, statins have been reported to inhibit LPA effects on RhoA activation [61], as well as the proliferation of smooth muscle cells and MCP-1 expression via Rac1 [62].

An additional possible source of serum ATX in disease states, beyond the adipose tissue, is the liver. Increased ATX expression has been reported in chronic liver diseases of

different etiologies, associated with shorter overall survival [27], while the genetic deletion of ATX from hepatocytes [27], or as discussed above adipocytes [51], attenuated liver steatosis and fibrosis. Therefore, increased levels of serum ATX are expected upon liver damage, whereas aberrant liver functions have been reported in COVID-19, irrespectively of pre-existing liver disease [63]. On the other hand, cirrhotic patients have high rates of liver failure and death from respiratory failure upon CoV-2 infection, attributed to increased systemic inflammation, immune dysfunction, and vasculopathy [63]. Therefore, ATX could be also a pathologic link between liver damage and COVID-19.

Plasma ATX levels have been recently reported to correlate with IL-6 levels in severe ARDS patients [47], as well as acute-on-chronic liver failure (ACLF) patients [64], as shown here in the serum of ICU COVID-19 patients (Figure 2). Increased serum IL-6 levels have been reported in COVID-19 patients, correlating with the severity of COVID-19 pneumonia and mortality risk [65], or respiratory failure and the need for mechanical ventilation [66]. Meta-analyses of published studies on COVID-19 laboratory findings indicated that serum levels of IL-6 were among the most predictive biomarkers [29,30]. Interestingly, components of the COVID-19 cytokine storm (IL-6, TNF, and IL-1 β) have been suggested to stimulate ATX expression and/or activity in different cell types, while, vice versa, LPA has been reported to stimulate TNF and IL-6 expression in different contexts [22], suggesting a possible interplay of the COVID-19 cytokine storm and the ATX/LPA axis.

Dex treatment, a potent suppressor of systemic inflammation including IL-6, has been shown to reduce mortality in hospitalized COVID-19 patients under oxygen supplementation treatment or mechanical ventilation [34,35]. Dex treatment has been shown to decrease ATX (as well as IL-6) levels in the mouse adipose tissue and plasma [67], as well as in irradiated mammary fat pad [68]. As shown here (Figure 3), Dex treatment of mechanically ventilated patients drastically reduced their ATX serum levels, indicating that the therapeutic effects of Dex in COVID-19 include the suppression of ATX serum levels.

An essential role for ATX/LPA in embryonic vasculature has been well established through genetic studies in both mice [69–71] and zebrafish [72]. In adult mice, *ENPP2* has been discovered as a high priority candidate gene for pulmonary hemorrhage upon SARS/MERS-CoV infection [73,74], while vascular leak has been suggested to be among the main pathological effects of ATX/LPA in pulmonary pathophysiology and fibrosis in mice [21,22]. As shown here, *ENPP2* mRNA expression in the COVID-19 lung tissue was detected mainly in artery cells D and Figure S3D), while high ATX expression from ECs in HEVs in lymph nodes has been previously reported [75]. Moreover, and in the same context, a plethora of LPA in vitro effects on endothelial cells has been suggested, with some controversy, including endothelial permeability, leukocyte adhesion, and cytokine expression, as previously reported in detail [26]. Among them, LPA has been suggested to stimulate the expression of E-sel from human aortic endothelial cells [76–78], a cell surface adhesion molecule regulating interaction with leukocytes. As shown here, ATX serum levels correlated with the corresponding sE-sel and sICAM serum levels (Figure S2), which has been independently associated, in the same samples, with mortality of COVID-19 ICU patients [33], suggesting that ATX/LPA effects in COVID-19 may also include vasculopathy.

IPF macrophages have been previously shown to stain for ATX, and conditional genetic deletion of ATX from macrophages (LysM⁺ cells) in mice, reduced BALF ATX levels and disease severity in modeled pulmonary fibrosis [20]. scRNAseq analysis of BALF cells from COVID-19 patients, where macrophages predominate, indicated that *ENPP2* mRNA expression was detected in different macrophage subpopulations (Figure 4C/UMAP, S3C/UMAP), pending FACS validation, where it could modulate their maturation in an autocrine mode via LPA [79–81]. LPA has also been suggested to stimulate, in vitro, the conversion of monocytes to DCs [38,82,83]. Interestingly, *ENPP2* mRNA expression was mainly detected in pDCs among all PBMCs and BALF cells in COVID-19 (Figure 4B,C and Figure S3B,C). pDCs are the principal interferon (IFN) type I producing cells in the human blood and can be rapidly recruited to inflamed sites [84]. Circulating

and lung pDCs have been shown to be diminished in COVID-19 [85,86], while IFN type I responses were highly impaired [87,88]. *ENPP2* mRNA expression was found upregulated in circulating pDCs (Figure 5B), and lung DCs (Figure 5E) from COVID-19 patients in comparison to cells from healthy controls. pDC development and homeostasis are regulated by the transcription TCF4 [89], which has been reported to be modulated by LPA in colon cancer cells [90], suggesting that *ENPP2* expression from pDCs and the local production of LPA modulates, in an autocrine manner, pDC development and homeostasis. The hypothesis is further supported from the genes that have been discovered, pending validation, to be increased in COVID-19 DCs, possibly regulated by *ENPP2* (Figure 5E). CD1a binds and presents to T-cell lipid metabolites and PLA2-synthesised fatty acid neoantigens and has been found to be expressed in immature DCs in mucosal surfaces, including the bronchus [39–41]. Tmem176B has also been associated with an immature state of dendritic cells [42,43], suggesting that *ENPP2* expression from COVID-19 pDCs, via LPA, delays their maturation. Although LPA signals in most cell types are considered pro-inflammatory and pro-surviving, an anti-inflammatory role of LPA, via LPAR2—the main subtype expressed in DCs (Figure S4B), has been proposed previously for DCs [38], further supporting a possible role for ATX/LPA in suppressing DC responses.

Taken together, a role for ATX/LPA in COVID-19 pathogenesis seems likely, possibly as a component of the cytokine storm perpetuating hyperinflammation and stimulating endothelial damage, as well as a regulator of the mononuclear phagocyte system and a suppressor of (p)DCs responses, non-withstanding its established role in fibrosis. Dex treatment in mechanically ventilated patients decreased ATX levels, indicating that the therapeutic effects of Dex in COVID-19 include the suppression of the ATX/LPA axis and that ATX levels can be druggable. More importantly, and given that COVID-19 and IPF share risk factors for disease severity, such as age/sex and comorbidities, existing and developing anti-fibrotic therapies have been suggested as additional therapeutic opportunities in COVID-19 [91–93]. One of the novel candidates target ATX is currently in clinical trials phase III in IPF [25]. Given the multiple possible roles of ATX in COVID-19, ATX inhibition could offer additional therapeutic options in COVID-19 management, both during and after hospitalization.

4. Materials and Methods

4.1. Human Patients and Samples

All studies were performed in accordance with the Helsinki Declaration principles. All collected data were anonymized in standardized forms, and informed consent was obtained from all individuals or patients' next-of-kin for severe cases. All available patient personal, epidemiological, clinical, and experimental data are summarized in the corresponding cohorts (Tables 1–3). All three cohorts were completely independent, and there was no overlap between swab and blood samples.

Cohort 1: Nasopharyngeal swab (NS) samples were collected upon routine diagnosis from adult patients tested positive in SARS-CoV2 RNA PCR and showing no or mild COVID-19 clinical symptoms, including cough, sore throat, mild fever below 38 °C, and loss of smell (positive mild group) or being hospitalized in the intensive care unit (ICU) with severe/critical symptoms, such as respiratory failure, septic shock, acute thrombosis, and multiorgan dysfunction (positive severe/critical group). The control group consisted of individuals with a negative SARS-CoV2 RNA PCR.

Cohort 2: Serum samples were collected with standardized procedures from COVID-19 patients admitted to the specialized COVID-19 WARD or to the intensive care unit (ICU) of the Evangelismos General Hospital from 24 March to 2 November 2020. SARS-CoV-2 infection was diagnosed by real-time reverse transcription PCR in nasopharyngeal swabs. The study was approved by the Evangelismos Hospital Research Ethics Committee (#170/24-4-2020).

Cohort 3: Serum samples were collected with standardized procedures from COVID-19 patients admitted to intensive care unit (ICU) of the University Hospital of Patras from

24 April to 6 December 2020. The study was approved by the University Hospital of Patras Research Ethics Committee (#216/08-05-2020).

4.2. Enzyme-Linked Immunosorbent Assay (ELISA)

ATX and IL-6 protein levels were quantified with dedicated ELISA kits according to the manufacturer's instructions (R&D Systems Inc., Minneapolis, MN, USA, and Invitrogen, ThermoFisher Scientific, Waltham, MA, USA, respectively). Measurements were performed in a blinded fashion in triplicate using a Triturus automated analyzer (Grifols, Barcelona, Spain). The presented results on ELISA quantification of soluble E-selectin (sE-sel) and P-selectin (sP-sel), ICAM, and ANG2 in the same samples, has been reported previously [33].

4.3. RNA Extraction and Q-RT-PCR

Total RNA extraction from nasopharyngeal swab samples was performed using a MagNA Pure LC Total Nucleic Acid Isolation Kit using a MagNA Pure LC 2.0 automated nucleic acid purifier (Roche, Basel, Switzerland), and viral RNA was quantified with a LightMix Modular SARS-CoV-2 RdRP Kit and a LightCycler Multiplex RNA Virus Master kit (Roche, Basel, Switzerland). *ENPP2* mRNA levels were quantified with Q-RT-PCR using a SYBR Green-based Luna[®] Universal qPCR Master Mix (New England Biolabs, Ipswich, MA, USA) (*ENPP2*: forward: 5'- ACT CAT GAA GAT GCA CAC AGC -3'; reverse 5'- CGC TCT CAT ATG TAT GCA GG -3'; product length 131 bp). Normalization was performed with the house-keeping gene 14-3-3-zeta polypeptide (YWHAZ), and the relative quantification method $2^{-\Delta\Delta C_t}$ was utilized.

4.4. Bulk/Single Cell RNA-seq Data Analysis and Mining

The available single cell RNA-seq object was mined for each one of the datasets (Table 1) using Seurat package v3 [94]. Marker selection and DEA were performed using Wilcoxon Rank Sum test ($FC > 1.2$; Bonferroni adj. $p < 0.05$). For identifying pDCs in the lung dataset of [13], DCs—as initially marked—were isolated, and principal components were calculated post to variable genes identification and data scaling using default parameters. The 30 first principal components were used to construct an SNN graph, while clusters were defined with a resolution of 0.8. pDCs were identified using marker genes reported in the cell atlas of [95].

Preprocessed read count matrices of [36] found here were analyzed using the metaseq-R2 package [96]. More specifically, reads were EDASeq normalized, filtered using default parameters, and then the PANDORA algorithm was used to combine the results of DESeq [97], DESeq2 [98], limma-voom [99], edgeR [100], and ABSSeq [101] methods. DEGs were defined using a $FC > 1.2$ and FDR adj. meta p -value < 0.05 .

4.5. Statistical Analysis

Statistical significance was assessed with Prism (GraphPad, San Diego, CA, USA) software with the appropriate test according to the distribution of values and their complexity, as detailed in each figure legend. Statistical tests used include the non-parametric Mann–Whitney U test, unpaired t -test, Spearman correlation, 2-way ANOVA followed by Bonferroni post hoc correction, Wilcoxon rank sum test-Bonferroni correction, and Kruskal–Wallis and Dunn post hoc test.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms221810006/s1>.

Author Contributions: Conceptualization, V.A.; Data curation, I.N., V.D., A.G.V., E.J., S.T., I.D., K.A. and A.T.; Formal analysis, I.N., D.F., K.N., P.M., G.M., M.E., A.G.V., N.V. and V.A.; Funding acquisition, V.A.; Investigation, I.N., V.D., I.D., E.A., K.A., A.K. and V.A.; Methodology, I.N., G.M., M.E., A.G.V. and N.V.; Project administration, V.A.; Resources, S.E.O., I.D., E.A., K.A., A.T. and A.K.; Software, D.F. and P.M.; Supervision, P.M., S.E.O., I.D., E.A., K.A., N.V., A.T., A.K. and V.A.;

Visualization, D.F.; Writing—original draft, I.N., D.F., K.N. and N.V.; Writing—review and editing, V.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship, and Innovation, under the call Research—Create—Innovate (project code: T1EDK-0049; recipient VA). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Research Ethics Committees of Evangelismos Hospital (#170/24-4-2020) and University Hospital of Patras (#216/08-05-2020).

Informed Consent Statement: All collected data were anonymized in standardized forms, and informed consent was obtained from all individuals or patients' next-of-kin for severe cases.

Data Availability Statement: All utilized, publicly available, datasets are detailed at Supplementary Table S1.

Acknowledgments: We would like to thank Alexandros Zacharis and Nikolaos Athanasiou for sample collection.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Hu, B.; Guo, H.; Zhou, P.; Shi, Z.-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **2020**, *19*, 141–154. [[CrossRef](#)]
- Osuchowski, M.F.; Winkler, M.S.; Skirecki, T.; Cajander, S.; Shankar-Hari, M.; Lachmann, G.; Monneret, G.; Venet, F.; Bauer, M.; Brunkhorst, F.M.; et al. The COVID-19 puzzle: Deciphering pathophysiology and phenotypes of a new disease entity. *Lancet Respir. Med.* **2021**, *9*, 622–642. [[CrossRef](#)]
- Huertas, A.; Montani, D.; Savale, L.; Pichon, J.; Tu, L.; Parent, F.; Guignabert, C.; Humbert, M. Endothelial cell dysfunction: A major player in SARS-CoV-2 infection (COVID-19)? *Eur. Respir. J.* **2020**, *56*, 2001634. [[CrossRef](#)] [[PubMed](#)]
- Fajgenbaum, D.C.; June, C.H. Cytokine Storm. *N. Engl. J. Med.* **2020**, *383*, 2255–2273. [[CrossRef](#)]
- Matthay, M.A.; Zemans, R.L.; Zimmerman, G.A.; Arabi, Y.M.; Beitler, J.R.; Mercat, A.; Herridge, M.; Randolph, A.G.; Calfee, C.S. Acute respiratory distress syndrome. *Nat. Rev. Dis. Primers* **2019**, *5*, 18. [[CrossRef](#)]
- Burnham, E.L.; Janssen, W.J.; Riches, D.W.; Moss, M.; Downey, G.P. The fibroproliferative response in acute respiratory distress syndrome: Mechanisms and clinical significance. *Eur. Respir. J.* **2014**, *43*, 276–285. [[CrossRef](#)] [[PubMed](#)]
- Maiese, A.; Manetti, A.C.; La Russa, R.; Di Paolo, M.; Turillazzi, E.; Frati, P.; Fineschi, V. Autopsy findings in COVID-19-related deaths: A literature review. *Forensic. Sci. Med. Pathol.* **2020**, *17*, 279–296. [[CrossRef](#)] [[PubMed](#)]
- Mo, X.; Jian, W.; Su, Z.; Chen, M.; Peng, H.; Peng, P.; Lei, C.; Chen, R.; Zhong, N.; Li, S. Abnormal pulmonary function in COVID-19 patients at time of hospital discharge. *Eur. Respir. J.* **2020**, *55*, 2001217. [[CrossRef](#)]
- Frija-Masson, J.; Debray, M.P.; Gilbert, M.; Lescure, F.X.; Travert, F.; Borie, R.; Khalil, A.; Crestani, B.; d'Ortho, M.P.; Bancal, C. Functional characteristics of patients with SARS-CoV-2 pneumonia at 30 days post-infection. *Eur. Respir. J.* **2020**, *56*, 81. [[CrossRef](#)] [[PubMed](#)]
- Gassel, R.J.J.; Bels, J.L.M.; Raafs, A.; Bussel, B.C.T.; Poll, M.C.G.; Simons, S.O.; Meer, L.W.L.; Gietema, H.A.; Posthuma, R.; Santen, S. High Prevalence of Pulmonary Sequelae at 3 Months after Hospital Discharge in Mechanically Ventilated Survivors of COVID-19. *Am. J. Respir. Crit. Care Med.* **2021**, *203*, 371–374. [[CrossRef](#)]
- George, P.M.; Barratt, S.L.; Condliffe, R.; Desai, S.R.; Devaraj, A.; Forrest, I.; Gibbons, M.A.; Hart, N.; Jenkins, R.G.; McAuley, D.F.; et al. Respiratory follow-up of patients with COVID-19 pneumonia. *Thorax* **2020**, *75*, 1009–1016. [[CrossRef](#)]
- Myall, K.J.; Mukherjee, B.; Castanheira, A.M.; Lam, J.L.; Benedetti, G.; Mak, S.M.; Preston, R.; Thillai, M.; Dewar, A.; Molyneaux, P.L.; et al. Persistent Post-COVID-19 Interstitial Lung Disease. An Observational Study of Corticosteroid Treatment. *Ann. Am. Thorac. Soc.* **2021**, *18*, 799–806. [[CrossRef](#)]
- Bharat, A.; Querrey, M.; Markov, N.S.; Kim, S.; Kurihara, C.; Garza-Castillon, R.; Manerikar, A.; Shilatifard, A.; Tomic, R.; Politanska, Y.; et al. Lung transplantation for patients with severe COVID-19. *Sci. Transl. Med.* **2020**, *12*, eabe4282. [[CrossRef](#)]
- Wu, M.; Chen, Y.; Xia, H.; Wang, C.; Tan, C.Y.; Cai, X.; Liu, Y.; Ji, F.; Xiong, P.; Liu, R.; et al. Transcriptional and proteomic insights into the host response in fatal COVID-19 cases. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 28336–28343. [[CrossRef](#)]
- McDonald, L.T. Healing after COVID-19: Are survivors at risk for pulmonary fibrosis? *Am. J. Physiol. Lung Cell. Mol. Physiol.* **2021**, *320*, L257–L265. [[CrossRef](#)] [[PubMed](#)]
- Esposito, A.J.; Menon, A.A.; Ghosh, A.J.; Putman, R.K.; Fredenburgh, L.E.; El-Chemaly, S.Y.; Goldberg, H.J.; Baron, R.M.; Hunninghake, G.M.; Doyle, T.J. Increased Odds of Death for Patients with Interstitial Lung Disease and COVID-19: A Case-Control Study. *Am. J. Respir. Crit. Care Med.* **2020**, *202*, 1710–1713. [[CrossRef](#)]

17. Drake, T.M.; Docherty, A.B.; Harrison, E.M.; Quint, J.K.; Adamali, H.; Agnew, S.; Babu, S.; Barber, C.M.; Barratt, S.; Bendstrup, E.; et al. Outcome of Hospitalization for COVID-19 in Patients with Interstitial Lung Disease. An International Multicenter Study. *Am. J. Respir. Crit. Care Med.* **2020**, *202*, 1656–1665. [[CrossRef](#)] [[PubMed](#)]
18. Beltramo, G.; Cottenet, J.; Mariet, A.S.; Georges, M.; Piroth, L.; Tubert-Bitter, P.; Bonniaud, P.; Quantin, C. Chronic respiratory diseases are predictors of severe outcome in COVID-19 hospitalised patients: A nationwide study. *Eur. Respir. J.* **2021**, *3*, 74. [[CrossRef](#)]
19. Magkrioti, C.; Galaris, A.; Kanellopoulou, P.; Stylianaki, E.A.; Kaffe, E.; Aidinis, V. Autotaxin and chronic inflammatory diseases. *J. Autoimmun.* **2019**, *104*, 102327. [[CrossRef](#)] [[PubMed](#)]
20. Oikonomou, N.; Mouratis, M.A.; Tzouveleki, A.; Kaffe, E.; Valavanis, C.; Vilaras, G.; Karameris, A.; Prestwich, G.D.; Bouros, D.; Aidinis, V. Pulmonary autotaxin expression contributes to the pathogenesis of pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* **2012**, *47*, 566–574. [[CrossRef](#)]
21. Tager, A.M.; LaCamera, P.; Shea, B.S.; Campanella, G.S.; Selman, M.; Zhao, Z.; Polosukhin, V.; Wain, J.; Karimi-Shah, B.A.; Kim, N.D.; et al. The lysophosphatidic acid receptor LPA1 links pulmonary fibrosis to lung injury by mediating fibroblast recruitment and vascular leak. *Nat. Med.* **2008**, *14*, 45–54. [[CrossRef](#)]
22. Ninou, I.; Magkrioti, C.; Aidinis, V. Autotaxin in Pathophysiology and Pulmonary Fibrosis. *Front. Med.* **2018**, *5*, 180. [[CrossRef](#)]
23. Ninou, I.; Kaffe, E.; Muller, S.; Budd, D.C.; Stevenson, C.S.; Ullmer, C.; Aidinis, V. Pharmacologic targeting of the ATX/LPA axis attenuates bleomycin-induced pulmonary fibrosis. *Pulm. Pharmacol. Ther.* **2018**, *52*, 32–40. [[CrossRef](#)]
24. Tager, A.M. Autotaxin emerges as a therapeutic target for idiopathic pulmonary fibrosis: Limiting fibrosis by limiting lysophosphatidic acid synthesis. *Am. J. Respir. Cell Mol. Biol.* **2012**, *47*, 563–565. [[CrossRef](#)]
25. Maher, T.M.; Kreuter, M.; Lederer, D.J.; Brown, K.K.; Wuyts, W.; Verbruggen, N.; Stutvoet, S.; Fieuw, A.; Ford, P.; Abi-Saab, W.; et al. Rationale, design and objectives of two phase III, randomised, placebo-controlled studies of GLPG1690, a novel autotaxin inhibitor, in idiopathic pulmonary fibrosis (ISABELA 1 and 2). *BMJ Open Respir. Res.* **2019**, *6*, e000422. [[CrossRef](#)]
26. Magkrioti, C.; Aidinis, V. ATX and LPA signalling in lung pathophysiology. *World J. Respirol.* **2013**, *3*, 77–103. [[CrossRef](#)]
27. Kaffe, E.; Katsifa, A.; Xylourgidis, N.; Ninou, I.; Zannikou, M.; Harokopos, V.; Foka, P.; Dimitriadis, A.; Evangelou, K.; Moulas, A.N.; et al. Hepatocyte autotaxin expression promotes liver fibrosis and cancer. *Hepatology* **2017**, *65*, 1369–1383. [[CrossRef](#)] [[PubMed](#)]
28. Farquhar, M.J.; Humphreys, I.S.; Rudge, S.A.; Wilson, G.K.; Bhattacharya, B.; Ciaccia, M.; Hu, K.; Zhang, Q.; Maily, L.; Reynolds, G.M.; et al. Autotaxin-lysophosphatidic acid receptor signalling regulates hepatitis C virus replication. *J. Hepatol.* **2017**, *9*, 16. [[CrossRef](#)] [[PubMed](#)]
29. Zhang, Z.-L.; Hou, Y.-L.; Li, D.-T.; Li, F.-Z. Laboratory findings of COVID-19: A systematic review and meta-analysis. *Scand. J. Clin. Lab. Investig.* **2020**, *80*, 441–447. [[CrossRef](#)] [[PubMed](#)]
30. Ulhaq, Z.S.; Soraya, G.V. Interleukin-6 as a potential biomarker of COVID-19 progression. *Méd. Mal. Infect.* **2020**, *50*, 382–383. [[CrossRef](#)]
31. Sun, S.; Wang, R.; Song, J.; Guan, M.; Li, N.; Zhang, X.; Zhao, Z.; Zhang, J. Blocking gp130 signaling suppresses autotaxin expression in adipocytes and improves insulin sensitivity in diet-induced obesity. *J. Lipid Res.* **2017**, *58*, 2102–2113. [[CrossRef](#)] [[PubMed](#)]
32. Castelino, F.V.; Bain, G.; Pace, V.A.; Black, K.E.; George, L.; Probst, C.K.; Goulet, L.; Lafyatis, R.; Tager, A.M. An Auto-taxin/Lysophosphatidic Acid/Interleukin-6 Amplification Loop Drives Scleroderma Fibrosis. *Arthritis Rheumatol.* **2016**, *68*, 2964–2974. [[CrossRef](#)] [[PubMed](#)]
33. Vassiliou, A.G.; Keskinidou, C.; Jahaj, E.; Gallos, P.; Dimopoulou, I.; Kotanidou, A.; Orfanos, S.E. ICU Admission Levels of Endothelial Biomarkers as Predictors of Mortality in Critically Ill COVID-19 Patients. *Cells* **2021**, *10*, 186. [[CrossRef](#)] [[PubMed](#)]
34. Horby, P.; Lim, W.S.; Emberson, J.R.; Mafham, M.; Bell, J.L.; Linsell, L.; Staplin, N.; Brightling, C.; Ustianowski, A.; Elmahi, E.; et al. Dexamethasone in Hospitalized Patients with Covid-19. *N. Engl. J. Med.* **2021**, *384*, 693–704.
35. Sterne, J.A.C.; Murthy, S.; Diaz, J.V.; Slutsky, A.S.; Villar, J.; Angus, D.C.; Annane, D.; Azevedo, L.C.P.; Berwanger, O.; Cavalcanti, A.B.; et al. Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19: A Meta-analysis. *JAMA* **2020**, *324*, 1330–1341. [[PubMed](#)]
36. Ota, M.; Nagafuchi, Y.; Hatano, H.; Ishigaki, K.; Terao, C.; Takeshima, Y.; Yanaoka, H.; Kobayashi, S.; Okubo, M.; Shirai, H.; et al. Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell* **2021**, *184*, 3006–3021. [[CrossRef](#)] [[PubMed](#)]
37. Yung, Y.C.; Stoddard, N.C.; Chun, J. LPA receptor signaling: Pharmacology, physiology, and pathophysiology. *J. Lipid Res.* **2014**, *55*, 1192–1214. [[CrossRef](#)]
38. Emo, J.; Meednu, N.; Chapman, T.J.; Rezaee, F.; Balys, M.; Randall, T.; Rangasamy, T.; Georas, S.N. Lpa2 Is a Negative Regulator of Both Dendritic Cell Activation and Murine Models of Allergic Lung Inflammation. *J. Immunol.* **2012**, *188*, 3784–3790. [[CrossRef](#)] [[PubMed](#)]
39. Tazi, A.; Bouchonnet, F.; Grandsaigne, M.; Boumsell, L.; Hance, A.J.; Soler, P. Evidence that granulocyte macrophage-colony-stimulating factor regulates the distribution and differentiated state of dendritic cells/Langerhans cells in human lung and lung cancers. *J. Clin. Investig.* **1993**, *91*, 566–576. [[CrossRef](#)]
40. Zeng, Z.; Castaño, A.R.; Segelke, B.W.; Stura, E.A.; Peterson, P.A.; Wilson, I.A. Crystal structure of mouse CD1: An MHC-like fold with a large hydrophobic binding groove. *Science* **1997**, *277*, 339–345. [[CrossRef](#)]

41. Bourgeois, E.A.; Subramaniam, S.; Cheng, T.Y.; De Jong, A.; Layre, E.; Ly, D.; Salimi, M.; Legaspi, A.; Modlin, R.L.; Salio, M.; et al. Bee venom processes human skin lipids for presentation by CD1a. *J. Exp. Med.* **2015**, *212*, 149–163. [[CrossRef](#)] [[PubMed](#)]
42. Condamine, T.; Le Texier, L.; Howie, D.; Lavault, A.; Hill, M.; Halary, F.; Cobbold, S.; Waldmann, H.; Cuturi, M.C.; Chiffolleau, E. Tmem176B and Tmem176A are associated with the immature state of dendritic cells. *J. Leukoc. Biol.* **2010**, *88*, 507–515. [[CrossRef](#)] [[PubMed](#)]
43. Picotto, G.; Morse, L.R.; Nguyen, N.; Saltzman, J.; Battaglino, R. TMEM176A and TMEM176B Are Candidate Regulators of Inhibition of Dendritic Cell Maturation and Function after Chronic Spinal Cord Injury. *J. Neurotrauma* **2020**, *37*, 528–533. [[CrossRef](#)]
44. Kostadinova, L.; Shive, C.L.; Anthony, D.D. Elevated Autotaxin and LPA Levels During Chronic Viral Hepatitis and Hepatocellular Carcinoma Associate with Systemic Immune Activation. *Cancers* **2019**, *11*, 1867. [[CrossRef](#)] [[PubMed](#)]
45. Joshita, S.; Ichikawa, Y.; Umemura, T.; Usami, Y.; Sugiura, A.; Shibata, S.; Yamazaki, T.; Fujimori, N.; Komatsu, M.; Matsumoto, A.; et al. Serum autotaxin is a useful liver fibrosis marker in patients with chronic hepatitis B virus infection. *Hepatol. Res.* **2018**, *48*, 275–285. [[CrossRef](#)]
46. Barbayanni, E.; Kaffe, E.; Aidinis, V.; Kokotos, G. Autotaxin, a secreted lysophospholipase D, as a promising therapeutic target in chronic inflammation and cancer. *Prog. Lipid Res.* **2015**, *58*, 76–96. [[CrossRef](#)]
47. Gao, L.; Li, X.; Wang, H.; Liao, Y.; Zhou, Y.; Wang, K.; Hu, J.; Cheng, M.; Zeng, Z.; Wang, T.; et al. Autotaxin levels in serum and bronchoalveolar lavage fluid are associated with inflammatory and fibrotic biomarkers and the clinical outcome in patients with acute respiratory distress syndrome. *J. Intensive Care* **2021**, *9*, 44. [[CrossRef](#)]
48. Casari, I.; Manfredi, M.; Metharom, P.; Falasca, M. Dissecting lipid metabolism alterations in SARS-CoV-2. *Prog. Lipid Res.* **2021**, *82*, 101092. [[CrossRef](#)]
49. Smyth, S.S.; Kraemer, M.; Yang, L.; Van Hoose, P.; Morris, A.J. Roles for lysophosphatidic acid signaling in vascular development and disease. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **2020**, *1865*, 158734. [[CrossRef](#)]
50. Dusaulcy, R.; Rancoule, C.; Gres, S.; Wanecq, E.; Colom, A.; Guigne, C.; van Meeteren, L.A.; Moolenaar, W.H.; Valet, P.; Saulnier-Blache, J.S. Adipose-specific disruption of autotaxin enhances nutritional fattening and reduces plasma lysophosphatidic acid. *J. Lipid Res.* **2011**, *52*, 1247–1255. [[CrossRef](#)] [[PubMed](#)]
51. Brandon, J.A.; Kraemer, M.; Vandra, J.; Halder, S.; Ubele, M.; Morris, A.J.; Smyth, S.S. Adipose-derived autotaxin regulates inflammation and steatosis associated with diet-induced obesity. *PLoS ONE* **2019**, *14*, e0208099. [[CrossRef](#)]
52. D'Souza, K.; Nzirorera, C.; Cowie, A.M.; Varghese, G.P.; Trivedi, P.; Eichmann, T.O.; Biswas, D.; Touaibia, M.; Morris, A.J.; Aidinis, V.; et al. Autotaxin-LPA signaling contributes to obesity-induced insulin resistance in muscle and impairs mitochondrial metabolism. *J. Lipid Res.* **2018**, *59*, 1805–1817. [[CrossRef](#)]
53. Reeves, V.L.; Trybula, J.S.; Wills, R.C.; Goodpaster, B.H.; Dube, J.J.; Kienesberger, P.C.; Kershaw, E.E. Serum Autotaxin/ENPP2 correlates with insulin resistance in older humans with obesity. *Obesity* **2015**, *23*, 2371–2376. [[CrossRef](#)]
54. D'Souza, K.; Paramel, G.V.; Kienesberger, P.C. Lysophosphatidic Acid Signaling in Obesity and Insulin Resistance. *Nutrients* **2018**, *10*, 399. [[CrossRef](#)]
55. Liu, Y.; Pan, Y.; Yin, Y.; Chen, W.; Li, X. Association of dyslipidemia with the severity and mortality of coronavirus disease 2019 (COVID-19): A meta-analysis. *Viol. J.* **2021**, *18*, 157. [[CrossRef](#)]
56. Adam, S.; Ho, J.H.; Bashir, B.; Iqbal, Z.; Ferdousi, M.; Syed, A.A.; Soran, H. The impact of atherosclerotic cardiovascular disease, dyslipidaemia and lipid lowering therapy on Coronavirus disease 2019 outcomes: An examination of the available evidence. *Curr. Opin. Lipidol.* **2021**, *32*, 231–243. [[CrossRef](#)] [[PubMed](#)]
57. Ferrara, F.; Vitiello, A. The advantages of drug treatment with statins in patients with SARS-CoV-2 infection. *Wien. Klin. Wochenschr.* **2021**, *16*, 1–8. [[CrossRef](#)]
58. Subir, R.; Jagat, J.M.; Kalyan, K.G. Pros and cons for use of statins in people with coronavirus disease-19 (COVID-19). *Diabetes Metab. Syndr.* **2020**, *14*, 1225–1229. [[CrossRef](#)] [[PubMed](#)]
59. Mitacchione, G.; Schiavone, M.; Curnis, A.; Arca, M.; Antinori, S.; Gasperetti, A.; Mascioli, G.; Severino, P.; Sabato, F.; Caracciolo, M.M.; et al. Impact of prior statin use on clinical outcomes in COVID-19 patients: Data from tertiary referral hospitals during COVID-19 pandemic in Italy. *J. Clin. Lipidol.* **2021**, *15*, 68–78. [[CrossRef](#)]
60. Daniels, L.B.; Sitapati, A.M.; Zhang, J.; Zou, J.; Bui, Q.M.; Ren, J.; Longhurst, C.A.; Criqui, M.H.; Messer, K. Relation of Statin Use Prior to Admission to Severity and Recovery Among COVID-19 Inpatients. *Am. J. Cardiol.* **2020**, *136*, 149–155. [[CrossRef](#)] [[PubMed](#)]
61. Kusama, T.; Mukai, M.; Ayaki, M.; Imamura, F.; Tatsuta, M.; Matsumoto, Y.; Nakamura, H.; Inoue, M. Inhibition of lysophosphatidic acid-induced RhoA activation and tumor cell invasion by 3-hydroxy-3-methylglutaryl-coenzyme A reductase inhibitors. *Int. J. Oncol.* **2003**, *23*, 1173–1178. [[CrossRef](#)]
62. Kaneyuki, U.; Ueda, S.; Yamagishi, S.; Kato, S.; Fujimura, T.; Shibata, R.; Hayashida, A.; Yoshimura, J.; Kojiro, M.; Oshima, K.; et al. Pitavastatin inhibits lysophosphatidic acid-induced proliferation and monocyte chemoattractant protein-1 expression in aortic smooth muscle cells by suppressing Rac-1-mediated reactive oxygen species generation. *Vasc. Pharmacol.* **2007**, *46*, 286–292. [[CrossRef](#)] [[PubMed](#)]
63. Marjot, T.; Webb, G.J.; Barritt, A.S.; Moon, A.M.; Stamataki, Z.; Wong, V.W.; Barnes, E. COVID-19 and liver disease: Mechanistic and clinical perspectives. *Nat. Rev. Gastroenterol. Hepatol.* **2021**, *18*, 348–364. [[CrossRef](#)]

64. Trovato, F.M.; Zia, R.; Napoli, S.; Wolfer, K.; Huang, X.; Morgan, P.E.; Husbyn, H.; Elgosbi, M.; Lucangeli, M.; Miquel, R.; et al. Dysregulation of the LPC-ATX-LPA axis in ACLF is associated with mortality and systemic inflammation via LPA-dependent monocyte activation. *Hepatology* **2021**, *3*, 22. [[CrossRef](#)]
65. Guirao, J.J.; Cabrera, C.M.; Jiménez, N.; Rincón, L.; Urra, J.M. High serum IL-6 values increase the risk of mortality and the severity of pneumonia in patients diagnosed with COVID-19. *Mol. Immunol.* **2020**, *128*, 64–68. [[CrossRef](#)] [[PubMed](#)]
66. Herold, T.; Jurinovic, V.; Arnreich, C.; Lipworth, B.J.; Hellmuth, J.C.; von Bergwelt-Baildon, M.; Klein, M.; Weinberger, T. Elevated levels of IL-6 and CRP predict the need for mechanical ventilation in COVID-19. *J. Allergy Clin. Immunol.* **2020**, *146*, 128–136.e4. [[CrossRef](#)]
67. Meng, G.; Tang, X.; Yang, Z.; Zhao, Y.; Curtis, J.M.; McMullen, T.P.W.; Brindley, D.N. Dexamethasone decreases the autotaxin-lysophosphatidate-inflammatory axis in adipose tissue: Implications for the metabolic syndrome and breast cancer. *FASEB J.* **2018**, *13*, 1226. [[CrossRef](#)] [[PubMed](#)]
68. Meng, G.; Wuest, M.; Tang, X.; Dufour, J.; McMullen, T.P.W.; Wuest, F.; Murray, D.; Brindley, D.N. Dexamethasone Attenuates X-Ray-Induced Activation of the Autotaxin-Lysophosphatidate-Inflammatory Cycle in Breast Tissue and Subsequent Breast Fibrosis. *Cancers* **2020**, *12*, 999. [[CrossRef](#)] [[PubMed](#)]
69. Tanaka, M.; Okudaira, S.; Kishi, Y.; Ohkawa, R.; Iseki, S.; Ota, M.; Noji, S.; Yatomi, Y.; Aoki, J.; Arai, H. Autotaxin stabilizes blood vessels and is required for embryonic vasculature by producing lysophosphatidic acid. *J. Biol. Chem.* **2006**, *281*, 25822–25830. [[CrossRef](#)]
70. van Meeteren, L.A.; Ruurs, P.; Stortelers, C.; Bouwman, P.; van Rooijen, M.A.; Pradere, J.P.; Pettit, T.R.; Wakelam, M.J.; Saulnier-Blache, J.S.; Mummery, C.L.; et al. Autotaxin, a secreted lysophospholipase D, is essential for blood vessel formation during development. *Mol. Cell. Biol.* **2006**, *26*, 5015–5022. [[CrossRef](#)]
71. Fotopoulou, S.; Oikonomou, N.; Grigorieva, E.; Nikitopoulou, I.; Paparountas, T.; Thanassopoulou, A.; Zhao, Z.; Xu, Y.; Kontoyiannis, D.L.; Remboutsika, E.; et al. ATX expression and LPA signalling are vital for the development of the nervous system. *Dev. Biol.* **2010**, *339*, 451–464. [[CrossRef](#)]
72. Yukiura, H.; Hama, K.; Nakanaga, K.; Tanaka, M.; Asaoka, Y.; Okudaira, S.; Arima, N.; Inoue, A.; Hashimoto, T.; Arai, H.; et al. Autotaxin regulates vascular development via multiple lysophosphatidic acid (LPA) receptors in zebrafish. *J. Biol. Chem.* **2011**, *286*, 43972–43983. [[CrossRef](#)]
73. Gralinski, L.E.; Menachery, V.D.; Morgan, A.P.; Totura, A.L.; Beall, A.; Kocher, J.; Plante, J.; Harrison-Shostak, D.C.; Schafer, A.; Pardo-Manuel de Villena, F.; et al. Allelic Variation in the Toll-Like Receptor Adaptor Protein Ticam2 Contributes to SARS-Coronavirus Pathogenesis in Mice. *G3 Bethesda* **2017**, *7*, 1653–1663. [[CrossRef](#)] [[PubMed](#)]
74. Gralinski, L.E.; Ferris, M.T.; Aylor, D.L.; Whitmore, A.C.; Green, R.; Frieman, M.B.; Deming, D.; Menachery, V.D.; Miller, D.R.; Buus, R.J.; et al. Genome Wide Identification of SARS-CoV Susceptibility Loci Using the Collaborative Cross. *PLoS Genet.* **2015**, *11*, e1005504. [[CrossRef](#)] [[PubMed](#)]
75. Kanda, H.; Newton, R.; Klein, R.; Morita, Y.; Gunn, M.D.; Rosen, S.D. Autotaxin, an ectoenzyme that produces lysophosphatidic acid, promotes the entry of lymphocytes into secondary lymphoid organs. *Nat. Immunol.* **2008**, *9*, 415–423. [[CrossRef](#)]
76. Takeda, Y.; Matoba, K.; Kawanami, D.; Nagai, Y.; Akamine, T.; Ishizawa, S.; Kanazawa, Y.; Yokota, T.; Utsunomiya, K. ROCK2 Regulates Monocyte Migration and Cell to Cell Adhesion in Vascular Endothelial Cells. *Int. J. Mol. Sci.* **2019**, *20*, 1331. [[CrossRef](#)] [[PubMed](#)]
77. Rizza, C.; Leitinger, N.; Yue, J.; Fischer, D.J.; Wang, D.A.; Shih, P.T.; Lee, H.; Tigyi, G.; Berliner, J.A. Lysophosphatidic acid as a regulator of endothelial/leukocyte interaction. *Lab. Invest.* **1999**, *79*, 1227–1235. [[PubMed](#)]
78. Shlyonsky, V.; Naeije, R.; Mies, F. Possible role of lysophosphatidic acid in rat model of hypoxic pulmonary vascular remodeling. *Pulm. Circ.* **2014**, *4*, 471–481. [[CrossRef](#)] [[PubMed](#)]
79. Ray, R.; Rai, V. Lysophosphatidic acid converts monocytes into macrophages in both mice and humans. *Blood* **2017**, *129*, 1177–1183. [[CrossRef](#)]
80. Ninou, I.; Sevastou, I.; Magkrioti, C.; Kaffe, E.; Stamatakis, G.; Thivaios, S.; Panayotou, G.; Aoki, J.; Kollias, G.; Aidinis, V. Genetic deletion of Autotaxin from CD11b+ cells decreases the severity of experimental autoimmune encephalomyelitis. *PLoS ONE* **2020**, *15*, e0226050. [[CrossRef](#)]
81. Santos-Nogueira, E.; Lopez-Serrano, C.; Hernandez, J.; Lago, N.; Astudillo, A.M.; Balsinde, J.; Estivill-Torrus, G.; de Fonseca, F.R.; Chun, J.; Lopez-Vales, R. Activation of Lysophosphatidic Acid Receptor Type 1 Contributes to Pathophysiology of Spinal Cord Injury. *J. Neurosci.* **2015**, *35*, 10224–10235. [[CrossRef](#)]
82. Martino, A.; Volpe, E.; Baldini, P.M. The influence of lysophosphatidic acid on the immunophenotypic differentiation of human monocytes into dendritic cells. *Haematologica* **2006**, *91*, 1273–1274.
83. Chen, R.; Roman, J.; Guo, J.; West, E.; McDyer, J.; Williams, M.A.; Georas, S.N. Lysophosphatidic acid modulates the activation of human monocyte-derived dendritic cells. *Stem Cells Dev.* **2006**, *15*, 797–804. [[CrossRef](#)]
84. Siegal, F.P.; Kadowaki, N.; Shodell, M.; Fitzgerald-Bocarsly, P.A.; Shah, K.; Ho, S.; Antonenko, S.; Liu, Y.J. The nature of the principal type 1 interferon-producing cells in human blood. *Science* **1999**, *284*, 1835–1837. [[CrossRef](#)]
85. Saichi, M.; Ladjemi, M.Z.; Korniotis, S.; Rousseau, C.; Ait Hamou, Z.; Massenet-Regad, L.; Amblard, E.; Noel, F.; Marie, Y.; Bouteiller, D.; et al. Single-cell RNA sequencing of blood antigen-presenting cells in severe COVID-19 reveals multi-process defects in antiviral immunity. *Nat. Cell Biol.* **2021**, *23*, 538–551. [[CrossRef](#)]

86. Zhou, R.; To, K.K.; Wong, Y.C.; Liu, L.; Zhou, B.; Li, X.; Huang, H.; Mo, Y.; Luk, T.Y.; Lau, T.T.; et al. Acute SARS-CoV-2 Infection Impairs Dendritic Cell and T Cell Responses. *Immunity* **2020**, *53*, 864–877.e5. [[CrossRef](#)]
87. Sánchez-Cerrillo, I.; Landete, P.; Aldave, B.; Sánchez-Alonso, S.; Sánchez-Azofra, A.; Marcos-Jiménez, A.; Ávalos, E.; Alcaraz-Serna, A.; de Los Santos, I.; Mateu-Albero, T.; et al. COVID-19 severity associates with pulmonary redistribution of CD1c+ DCs and inflammatory transitional and nonclassical monocytes. *J. Clin. Investig.* **2020**, *130*, 6290–6300. [[CrossRef](#)] [[PubMed](#)]
88. Hadjadj, J.; Yatim, N.; Barnabei, L.; Corneau, A.; Boussier, J.; Smith, N.; Péré, H.; Charbit, B.; Bondet, V.; Chenevier-Gobeaux, C.; et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **2020**, *369*, 718–724. [[CrossRef](#)] [[PubMed](#)]
89. Cisse, B.; Caton, M.L.; Lehner, M.; Maeda, T.; Scheu, S.; Locksley, R.; Holmberg, D.; Zweier, C.; den Hollander, N.S.; Kant, S.G.; et al. Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. *Cell* **2008**, *135*, 37–48. [[CrossRef](#)] [[PubMed](#)]
90. Guo, L.; He, P.; No, Y.R.; Yun, C.C. Krüppel-like factor 5 incorporates into the β -catenin/TCF complex in response to LPA in colon cancer cells. *Cell. Signal.* **2015**, *27*, 961–968. [[CrossRef](#)] [[PubMed](#)]
91. Spagnolo, P.; Balestro, E.; Aliberti, S.; Cocconcelli, E.; Biondini, D.; Casa, G.D.; Sverzellati, N.; Maher, T.M. Pulmonary fibrosis secondary to COVID-19: A call to arms? *Lancet Respir. Med.* **2020**, *8*, 750–752. [[CrossRef](#)]
92. George, P.M.; Wells, A.U.; Jenkins, R.G. Pulmonary fibrosis and COVID-19: The potential role for antifibrotic therapy. *Lancet Respir. Med.* **2020**, *6*, 3022. [[CrossRef](#)]
93. Vasarmidi, E.; Tsitoura, E.; Spandidos, D.A.; Tzanakis, N.; Antoniou, K.M. Pulmonary fibrosis in the aftermath of the COVID-19 era (Review). *Exp. Ther. Med.* **2020**, *20*, 2557–2560. [[CrossRef](#)]
94. Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck, W.M., III; Hao, Y.; Stoerckius, M.; Smibert, P.; Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177*, 1888–1902.e21. [[CrossRef](#)]
95. Travaglini, K.J.; Nabhan, A.N.; Penland, L.; Sinha, R.; Gillich, A.; Sit, R.V.; Chang, S.; Conley, S.D.; Mori, Y.; Seita, J.; et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **2020**, *587*, 619–625. [[CrossRef](#)]
96. Fanidis, D.; Moulos, P. Integrative, normalization-insusceptible statistical analysis of RNA-Seq data, with improved differential expression and unbiased downstream functional analysis. *Brief. Bioinform.* **2020**, *22*, 156. [[CrossRef](#)]
97. Anders, S.; Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **2010**, *11*, R106. [[CrossRef](#)] [[PubMed](#)]
98. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
99. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [[CrossRef](#)]
100. McCarthy, D.J.; Chen, Y.; Smyth, G.K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **2012**, *40*, 4288–4297. [[CrossRef](#)] [[PubMed](#)]
101. Yang, W.; Rosenstiel, P.C.; Schulenburg, H. ABSSeq: A new RNA-Seq analysis method based on modelling absolute expression differences. *BMC Genom.* **2016**, *17*, 484. [[CrossRef](#)] [[PubMed](#)]

Lysophosphatidic Acid Is a Proinflammatory Stimulus of Renal Tubular Epithelial Cells

Το λυσοφωσφατιδικό οξύ είναι ένα προφλεγμονώδες ερέθισμα των επιθηλιακών κυττάρων εσπειραμένων σωληνοειδών

Ο όρος χρόνια νεφρική νόσος (CKD) αναφέρεται σε ένα φάσμα ασθενειών που χαρακτηρίζονται από νεφρική ίνωση, μόνιμες αλλοιώσεις της νεφρικής δομής και χαμηλό ρυθμό σπειραματικής διήθησης. Η παρατεταμένη καταστροφή του επιθηλίου των σωληνοειδών περιλαμβάνει μία σειρά από αλλαγές που εν τέλει οδηγούν σε CKD, υπογραμμίζοντας της σημασία του επιθηλίου των σωληνοειδών στην εν λόγω διαδικασία. Το λυσοφωσφατιδικό οξύ (LPA) είναι ένα βιοενεργό λιπίδιο που σηματοδοτεί κυρίως μέσω έξι συγγενών υποδοχέων και εμπλέκεται σε διάφορες παθολογικές συνθήκες χρόνιας φλεγμονής. Σε αυτή την αναφορά ερεθίσαμε ανθρώπινα επιθηλιακά κύτταρα των εγγύς εσπειραμένων σωληναρίων (HKC-8) με LPA και άλλους 175 πιθανούς παθολογικούς παράγοντες, ενώ ταυτοχρόνως ανιχνεύσαμε τα επίπεδα 27 ενδοκυττάρων φωσφοπρωτεϊνών και 32 εξωκυττάρως εκκρινόμενων μορίων με μέθοδο πολλαπλής ELISA. Αυτή η ποσοτικοποίηση αποκάλυψε μία μεγάλη ποσότητα πληροφοριών σχετικά με την σηματοδότηση και την φυσιολογία των HKC-8 κυττάρων, οι οποίες μπορούν να επεκταθούν και σε άλλα επιθηλιακά κύτταρα των εγγύς εσπειραμένων σωληναρίων. Οι αποκρίσεις στο LPA ομαδοποιήθηκαν μαζί με προφλεγμονώδη ερεθίσματα, όπως αυτό του TNF και της IL-1, προωθώντας τη φωσφορυλίωση σημαντικών φλεγμονωδών κόμβων σηματοδότησης, συμπεριλαμβανομένων των CREB1, ERK1, JUN, IκBa και MEK1, όπως επίσης και την έκκριση φλεγμονωδών παραγόντων κλινικής σημασίας, συμπεριλαμβανομένων των CCL2, CCL3, CXCL10, ICAM1, IL-6, and IL-8. Τα περισσότερα εξ αυτών παρατηρήθηκαν για πρώτη φορά στα επιθηλιακά κύτταρα των εγγύς εσπειραμένων σωληναρίων. Η επαγωγή μονοπατιών μεταγωγής σήματος από το LPA επιβεβαιώθηκε με την χρήση φαρμακολογικών δοκιμασιών. Η ανίχνευση των παραπάνω μονοπατιών και η έκκριση φλεγμονωδών παραγόντων προσφέρουν καινοτόμα στοιχεία για τον πιθανό ρόλο του LPA στην παθογένεση της CKD.



Article

Lysophosphatidic Acid Is a Proinflammatory Stimulus of Renal Tubular Epithelial Cells

Christiana Magkrioti ^{1,†}, Georgia Antonopoulou ^{1,†}, Dionysios Fanidis ¹, Vaia Pliaka ², Theodore Sakellaropoulos ², Leonidas G. Alexopoulos ^{2,3}, Christoph Ullmer ⁴ and Vassilis Aidinis ^{1,*}

¹ Institute for Fundamental Biomedical Research, Biomedical Sciences Research Center Alexander Fleming, 16672 Athens, Greece; magkrioti@fleming.gr (C.M.); antonopoulou@fleming.gr (G.A.); fanidis@fleming.gr (D.F.)

² ProtATonce Ltd., 15343 Athens, Greece; vicky.pliaka@protatonce.com (V.P.); teosakel@gmail.com (T.S.); leo@protatonce.com (L.G.A.)

³ School of Mechanical Engineering, National Technical University of Athens, 15780 Zografou, Greece

⁴ Roche Pharmaceutical Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd., 4070 Basel, Switzerland; christoph.ullmer@roche.com

* Correspondence: aidinis@fleming.gr

† These authors contributed equally to this work.



Citation: Magkrioti, C.; Antonopoulou, G.; Fanidis, D.; Pliaka, V.; Sakellaropoulos, T.; Alexopoulos, L.G.; Ullmer, C.; Aidinis, V. Lysophosphatidic Acid Is a Proinflammatory Stimulus of Renal Tubular Epithelial Cells. *Int. J. Mol. Sci.* **2022**, *23*, 7452. <https://doi.org/10.3390/ijms23137452>

Academic Editor: Dong Soon Im

Received: 10 June 2022

Accepted: 2 July 2022

Published: 5 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Chronic kidney disease (CKD) refers to a spectrum of diseases defined by renal fibrosis, permanent alterations in kidney structure, and low glomerular-filtration rate. Prolonged epithelial-tubular damage involves a series of changes that eventually lead to CKD, highlighting the importance of tubular epithelial cells in this process. Lysophosphatidic acid (LPA) is a bioactive lipid that signals mainly through its six cognate LPA receptors and is implicated in several chronic inflammatory pathological conditions. In this report, we have stimulated human proximal tubular epithelial cells (HKC-8) with LPA and 175 other possibly pathological stimuli, and simultaneously detected the levels of 27 intracellular phosphoproteins and 32 extracellular secreted molecules with multiplex ELISA. This quantification revealed a large amount of information concerning the signaling and the physiology of HKC-8 cells that can be extrapolated to other proximal tubular epithelial cells. LPA responses clustered with pro-inflammatory stimuli such as TNF and IL-1, promoting the phosphorylation of important inflammatory signaling hubs, including CREB1, ERK1, JUN, IκBα, and MEK1, as well as the secretion of inflammatory factors of clinical relevance, including CCL2, CCL3, CXCL10, ICAM1, IL-6, and IL-8, most of them shown for the first time in proximal tubular epithelial cells. The identified LPA-induced signal-transduction pathways, which were pharmacologically validated, and the secretion of the inflammatory factors offer novel insights into the possible role of LPA in CKD pathogenesis.

Keywords: lysophosphatidic acid; inflammation; tubular epithelial cells; cytokines

1. Introduction

Chronic kidney disease (CKD), with a worldwide prevalence of 13.4% [1], refers to a spectrum of diseases defined by permanent alterations in kidney structure or function. The most prominent pathological characteristic of CKD is renal fibrosis, while the gradual replacement of podocytes (in the glomeruli) and the tubulointerstitium with an extracellular matrix (ECM) leads to irreversible nephron loss [2,3]. Renal tubular epithelial cells (TECs) are the cells lining the nephrons, and the ones responsible for the selective transport of minerals, organic compounds, and water into and out of the tubular fluid of the nephrons, however, they are vulnerable to injuries. An injury can cause the loss of epithelial-cell polarization and intercellular contacts, epithelial to mesenchymal transition (EMT), cell death and, finally, abrogation of the tubular function [3]. The initial injury is followed by repair mechanisms and epithelial regeneration; but, depending on the severity and frequency of

the initial injuries, these repair mechanisms may become maladaptive and the epithelial damage may progress to CKD [3–5]. Prolonged epithelial-tubular damage involves cell death, mitochondrial dysfunction, metabolic disturbance, oxidative stress, cell-cycle arrest and senescence, partial EMT, epigenomic modifications, and proinflammatory mediators' secretion [3,5–8], with the latter further fueling the tubular injury [9]. Damaged TECs express CCL2/ MCP1 (monocyte chemoattractant protein-1) and CCL5/RANTES (regulated on activation, normal T cell expressed and secreted), thus mobilizing macrophages and dendritic cells to the site of the injury, a factor associated with CKD [10]. Macrophages, on their part, secrete a number of factors that further activate inflammation and fibrosis, although their functions vary depending on their polarization [8]. Furthermore, during the transition to CKD and the phenotypical changes of epithelial cells, the tubular basement membrane may be disrupted by activated matrix metalloproteinases allowing myofibroblast intrusion in the interstitium and, thus, promoting tubulointerstitial fibrosis and CKD progression [3]. All the above highlight the importance of TECs as initiators of tubulointerstitial fibrosis [8].

Apart from fibrosis, the progression of CKD has been shown to lead to inflammation and oxidative stress, suggesting that CKD is a low-grade inflammatory process [11,12]. Indeed, the lower GFR and the higher albuminuria of CKD patients are associated with higher levels of inflammatory cytokines, such as IL-1 β and IL-6 in their plasma [13]. This upsurge is partly owed to the prolonged half-life of cytokines due to the impaired excretory renal function but also to increased tissue production. Subsequently, this amplified chronic inflammation may lead to high mortality in CKD patients [13].

Lysophosphatidic acid (LPA) is a lysophospholipid present in most biological fluids. LPA is actually a mixture of species carrying various saturated or unsaturated fatty acids. LPA presents many functions due to its signaling through at least six cognate receptors (LPAR1-6), which are further coupled with G proteins, activating numerous signal-transduction pathways [14]. LPA signaling is implicated in several chronic inflammatory or fibrotic diseases, such as rheumatoid arthritis (RA), cardiovascular diseases, pulmonary and liver fibrosis, and others [14]. Additionally, LPA and its receptors are also involved in CKD. As early as 1998, LPA levels in the plasma of patients with renal failure on hemodialysis were found to be higher compared to healthy controls [15]. Two LPA species (16:0 and 18:2) were among a panel of seven metabolites that discriminated the sera of patients with CKD of diverse aetiologies from the sera of healthy subjects [16]. Moreover, LPA 16:0 and LPA 20:4 were found to have risen in the urine of type II diabetes patients with nephropathy compared to type II diabetes patients without nephropathy, probably due to higher local production [17]. Furthermore, the LPA-producing enzyme, Autotaxin (ATX), and LPAR1 and LPAR3 were found to have increased in diabetic human kidneys compared to healthy kidneys [18]. Concerning TECs, LPA promotes the expression or activation of profibrotic molecules [19]. Specifically, LPA induces integrin α v β 6-mediated latent transforming growth factor beta (TGF β) activation, which leads to the increase in connective tissue growth factor (CTGF) and platelet-derived growth factor (PDGFB) mRNA and protein expression in a TGF β -dependent fashion in several TEC cell lines and primary cultures [19]. Additionally, LPA treatment on mouse renal TECs induces TGF β mRNA expression [20,21]. The above conclude that LPA has an effect in the secretory pattern of fibrosis-related factors from TECs.

In an effort to expand our knowledge on TECs' signaling and secretion of proinflammatory/profibrotic factors, we exposed the kidney proximal tubular epithelial cell line HKC-8, which derives from the normal kidney cortex [22], to 175 inflammatory-immunological stimuli and measured the levels of 27 intracellular phosphoproteins as well as 32 extracellular secreted mediators upon each one of the stimuli employing custom multiplex ELISA. Furthermore, we exposed HKC-8 to three different species of LPA in order to investigate the effect of LPA on TECs. According to our findings, LPA induced the phosphorylation of CREB1, ERK1, I κ B α , JUN, and MEK1 and the secretion of proinflammatory molecules CCL2, CCL3, CXCL10, ICAM1, IL-6, and IL-8.

2. Results

2.1. Exposure of HKC-8 Cells to LPA and 175 Other Stimuli

In an attempt to shed light on the signaling of TECs, we employed the human renal proximal tubular epithelial cell line HKC-8 and exposed it to LPA and 175 other inflammatory-immunological stimuli. Stimuli comprised toll-like receptor (TLR) ligands, cytokines, chemokines, growth factors and drugs (Tables S1 and S2). The experimentation included two sub-experiments. In the first sub-experiment, cells were exposed to the stimuli for 24 h and supernatants were collected, while in the second sub-experiment the exposure lasted for 5 or 25 min and cell lysates were collected. Subsequently, we measured the levels of 32 extracellular secreted factors in the supernatants and the levels of 27 intracellular phosphoproteins in the lysates upon each one of the stimuli employing custom multiplex ELISA. This method is based on the usage of magnetic beads of unique spectral signatures, conjugated with antibodies against each of the analytes. Recognition is achieved with biotinylated detection antibodies and streptavidin conjugated with R-Phycoerythrin (Figure 1). For each stimulus, the levels of the analytes at the stimulated state were divided with the levels of the analytes at the unstimulated state (median of the control wells) and the emerging ratios (fold-changes, FCs) correspond to the normalized responses. The responses to the different stimuli are depicted as FCs in the heatmap plots of Figures 2 and 3. These plots show in detail the activations (or not) of each analyte upon each stimulus. The color code indicates the level of response, with light blue referring to low or no response and dark red referring to a high response/ activation. A response was considered active when the expression of an analyte upon a stimulus was higher than 1.5 FCs, which was set as the threshold in our analysis. The selection of the threshold was made based on the sensitivity analysis of Supplementary Figure S1, where it can be seen that at a threshold equal to 1.5, the number of activations is rather stable and also high enough to allow for the subsequent analysis. Regarding the phosphoproteins' sub-experiment, which was done at two time points (5' and 25'), the heatmap depicts an active response when the FC is above the threshold at at least one time point.

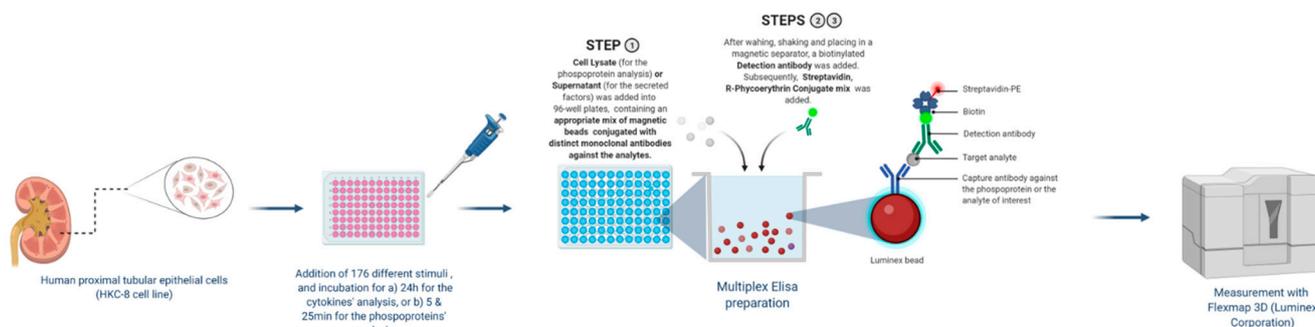


Figure 1. Multiplex ELISA for the detection of secreted factors and signaling molecules phosphorylation. Human kidney proximal tubular epithelial cells, HKC-8, were stimulated with 176 stimuli. Supernatants were collected at 24 h and cell lysates at 5 and 25 min post stimulation. Supernatants or cell lysates were added to a mix containing magnetic beads internally dyed with precise proportions of red and infrared fluorophores, thus, rendering unique spectral signature microspheres. Each unique microsphere-bead was conjugated with a distinct monoclonal antibody against a secreted factor or a phosphoprotein. Biotinylated detection antibodies were added to the mix, followed by a streptavidin-R-Phycoerythrin complex. This process allows the simultaneous recognition of 32 secreted factors or 27 phosphoproteins in one sample. Created with [BioRender.com](https://www.biorender.com), accessed on 1 June 2022.

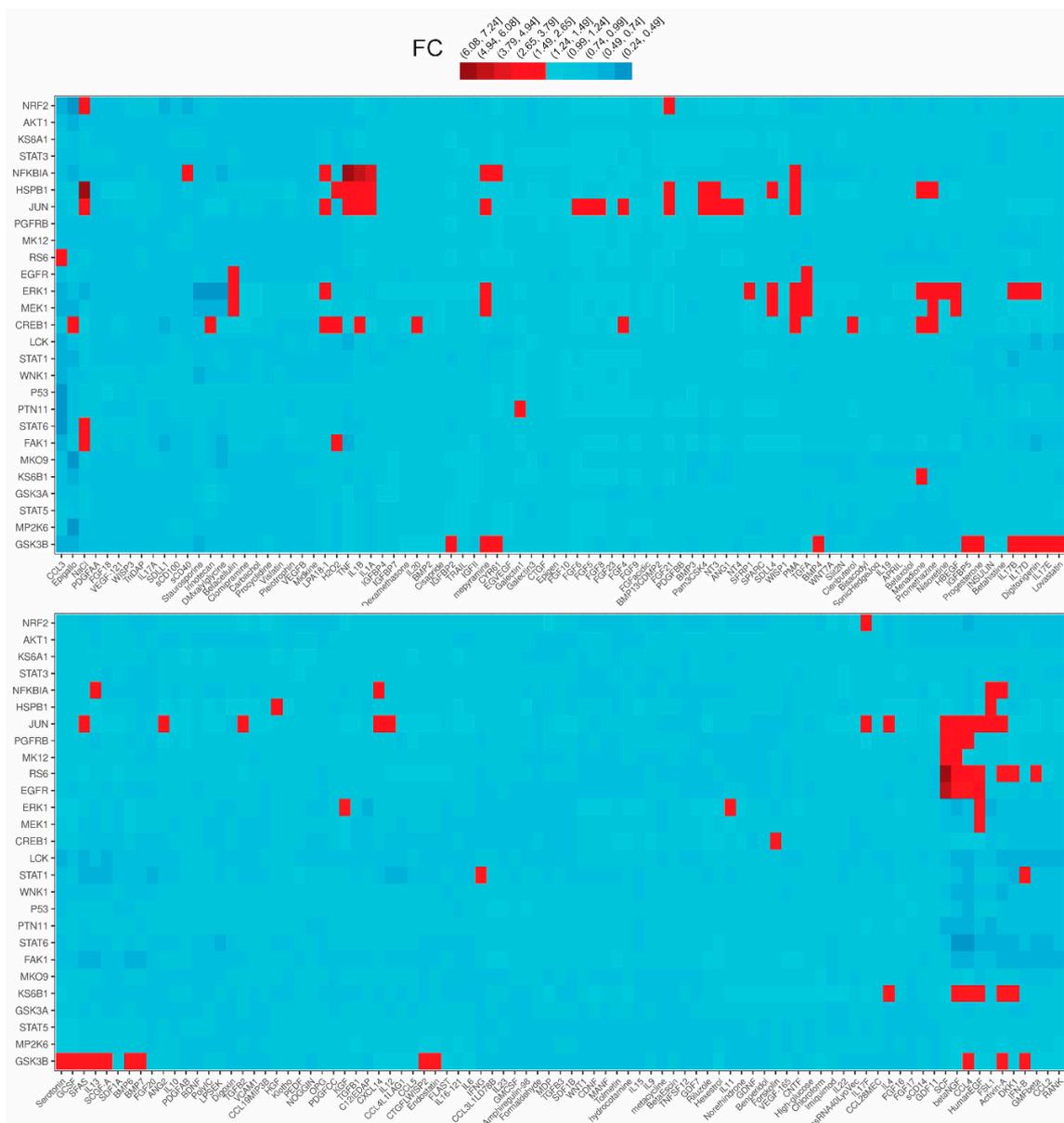


Figure 3. Phosphorylation of 27 major hubs in intracellular signaling pathways of human kidney proximal tubular epithelial cells (HKC-8) upon the stimulation with LPA (18:1) and 175 disparate biological stimuli. The expression was assessed with multiplex ELISA employing microbeads of unique spectral signatures conjugated with monoclonal antibodies specific for each of the 27 phosphoproteins. Red indicates active signals ($FC \geq 1.5$). See also Figure S1.

Among the 176 stimuli tested, 108 were the active ones at either sub-experiment. In total, 76 stimuli gave an active response in the secreted factors sub-experiment (Figure 2) and 71 in the phosphoproteins sub-experiment (Figure 3). Collectively, hepatocyte growth factor (HGF) was the stimulus that evoked the most active responses (18 secreted factors, 0 phosphoproteins), followed by interleukin 1β (13 secreted factors, 4 phosphoproteins), the synthetic TLR2/TLR6 ligand FSL1 (10 secreted factors, 3 phosphoproteins), Tumor-necrosis factor alpha or $TNF\alpha$ (9 secreted factors, 3 phosphoproteins), phorbol 12-myristate 13-acetate or PMA (6 secreted factors, 6 phosphoproteins), angiopoietin or ANG1 (11 secreted factors), digoxin (11 secreted factors), and others. Among the secreted factors, chemokine CXCL10 was the most common active responder (34 hits), followed by CCL3 (30 hits), while, among the signaling phosphoproteins, JUN was the molecule with the most activations (29 hits).

2.2. LPA Is a Proinflammatory Stimulus

One of the 176 stimuli to which HKC-8 cells were exposed was LPA 18:1, which led to several activations. Concerning the secreted factors, it induced the secretion of interleukin 6 (IL-6), interleukin 8 (IL-8), and chemokines CCL2, CCL3, and CXCL10 (Figure 2). With regard to the phosphoproteins, LPA 18:1 induced the phosphorylation of JUN, I κ B α , ERK1, and CREB1 (Figure 3). Subsequently, in an effort to verify the LPA results, we performed a second experiment using two more LPA species, LPA 16:0 and 20:4, on top of LPA 18:1. This experiment was done in triplicates, thus allowing statistical analysis. With a threshold of FC \geq 1.5, a number of active signals was detected upon the three LPA species (Figures 4 and 5). Most of the responses were shared between them, but some differences exist, which, however, need further exploration. Regarding the secreted factors, CCL3, IL-6, and IL-8 were expressed statistically significantly higher than the medium upon all three species (Figure 4). CCL2 was expressed at levels higher than the medium upon them as well, however, not statistically significantly. Soluble Intercellular Adhesion Molecule 1 (sICAM1) was also induced upon all three species statistically significantly, however, only upon LPA 16:0 and 20:4 above the FC = 1.5 (Figure 4). LPA 18:1 also triggered the expression of CXCL10 statistically significantly.

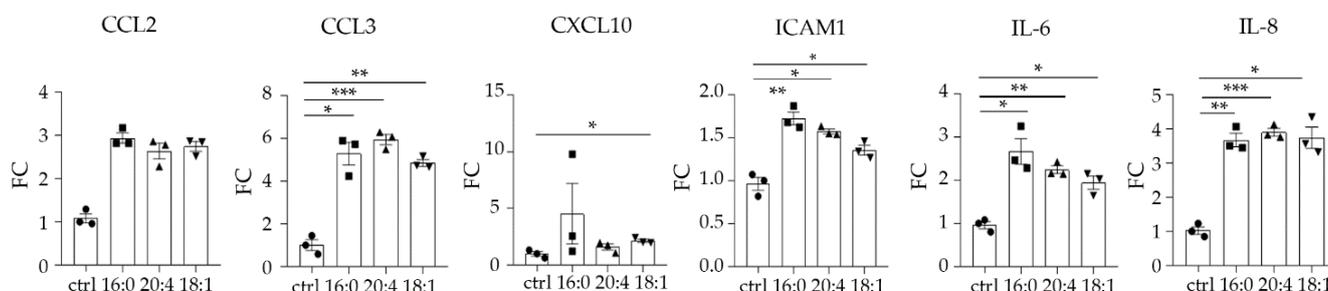


Figure 4. LPA stimulates the secretion of CCL2, CCL3, CXCL10, ICAM1, IL-6, and IL-8 from human kidney proximal tubular epithelial cells (HKC-8). Multiplex ELISA quantifying the expression of the indicated secreted factors in the supernatants from HKC-8 cells upon the stimulation with three different LPA species (16:0, 20:4, 18:1) at 10 μ M for 24 h. Statistical significance was assessed with Brown–Forsythe’s and Welch’s ANOVA followed by Dunnett’s post hoc test in the case of normal distribution or with Kruskal–Wallis test in the case of non-normal distribution; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Circles correspond to control values, squares correspond to LPA 16:0 values, upward triangles correspond to LPA 20:4 values and downward triangles correspond to LPA 18:1 values.

Pertaining to the signaling phosphorylations, components that were phosphorylated above the threshold upon the three LPA species were I κ B α (25’) and CREB1 (25’), with a statistical significance for LPA 20:4 (Figure 5). JUN (25’) was phosphorylated by all three species with statistically significant phosphorylation upon LPA 16:0 (although below the FC = 1.5) and 20:4. JUN (25’) gave an active signal upon LPA 18:1 too, albeit not statistically significantly. Moreover, LPA 18:1 statistically significantly induced the phosphorylation of MEK1 (5’) and CREB1 (5’) (Figure 5). Finally, ERK1 (5’) was phosphorylated upon LPA 18:1 and ERK1 (25’) upon LPA 20:4, but not statistically significantly.

2.3. Clustering and Responses of Other Stimuli

Trying to identify stimuli that affect the same kidney processes, we initially transformed absolute Luminex data into fold-change values, with respect to the median value of medium-treated cells, separately for secreted factors (Figure 2) and phosphoproteins plates (Figure 3). Subsequently, we concatenated all plates and focused on 108 stimuli that caused at least one activation event (activation threshold FC \geq 1.5). After removing non-responsive signals, 46 signals remained. Their respective values were binary transformed (1: activation; 0: non-activation) and used for stimuli unsupervised clustering (Gower’s metric; divisive clustering), which led to the definition of three clusters (Figure 6).

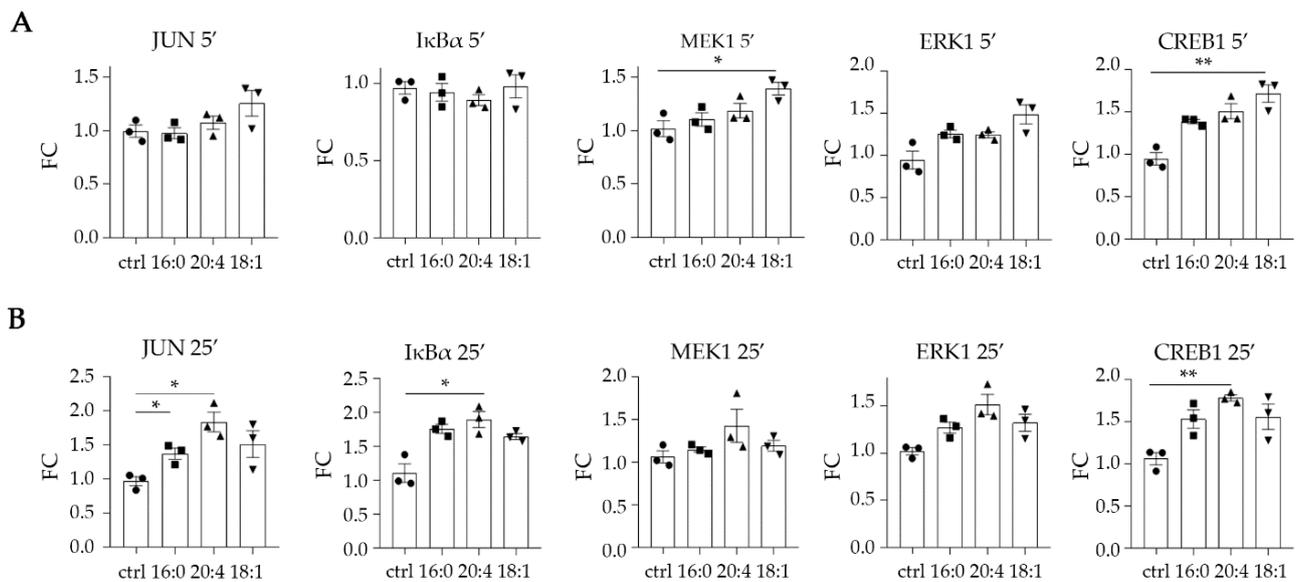


Figure 5. LPA stimulates the phosphorylation of JUN, IκBα, MEK1, ERK1, and CREB1 in human kidney proximal tubular epithelial cells (HKC-8). Cells were incubated with three different LPA species (16:0, 20:4, 18:1) at 10 μM for 5 (A) or 25 min (B), and the phosphorylation was assessed with multiplex ELISA in triplicates. Circles correspond to control values, squares correspond to LPA 16:0 values, upward triangles correspond to LPA 20:4 values and downward triangles correspond to LPA 18:1 values. Statistical significance was assessed with Brown–Forsythe’s and Welch’s ANOVA followed by Dunnett’s post hoc test in the case of normal distribution or with Kruskal–Wallis test in the case of non-normal distribution. * $p < 0.05$, ** $p < 0.01$.

As shown in Figure 6, LPA falls into the same cluster (cluster 2) as PMA, which is a PKC activator and a T-cell activator; IL-1α, IL-1β, IL-17α, and TNFα, which are proinflammatory cytokines; FSL1, a lipoprotein derived from *Mycoplasma salivarium* and a TLR ligand; PolyIC, a synthetic analogue of double-stranded RNA and a TLR3 ligand; promethazine, an antihistamine drug; TNFSF12/TWEAK, which is a TNF family member; and CXCL14/BRAK that is a breast and kidney-expressed chemokine activating B cells and monocytes. Between LPA and the aforementioned molecules of its cluster, many similarities appear in the secreted factors pattern induction; IL-6, IL-8 and CCL3 are induced by all 11 members of the cluster, while CCL2 and CXCL10 are induced by most of them. The results of the phosphoproteins also indicate the signaling convergence between LPA and the other molecules of cluster 2, as most of them induce the phosphorylation of JUN and IκBα. Promethazine, PMA and IL-1β also lead to the phosphorylation of CREB1, like LPA does.

Another cluster that emerges is the cluster 1, which includes angiotensin 1 (ANG1), HGF, platelet-derived growth factor AB (PDGFAB), PDGFBB, digoxin, bone morphogenetic protein 2 (BMP2), brain-derived neurotrophic factor (BDNF), and IL-10. These stimuli induce a series of secreted cytokines such as Prokineticin 1 (PROK1), Ciliary Neurotrophic Factor (CNTF), TNF family members, IL-3, CXCL11, and others. ANG1, which triggered the secretion of several factors (IL-3, TNFα, TNF10, PROK1, CNTF, TNF12, VEGFB, CXCL11, IL-5, CXCL10, CCL5, IL-3), is thought to be protective in models of renal injury [23], while BMP2, which induced TSLP, IL-3, TNFα, TNF10, PROK1, and IL-5, has been found to induce the commitment of adult renal progenitor cells (ARPCs) toward a myofibroblastic phenotype [24].

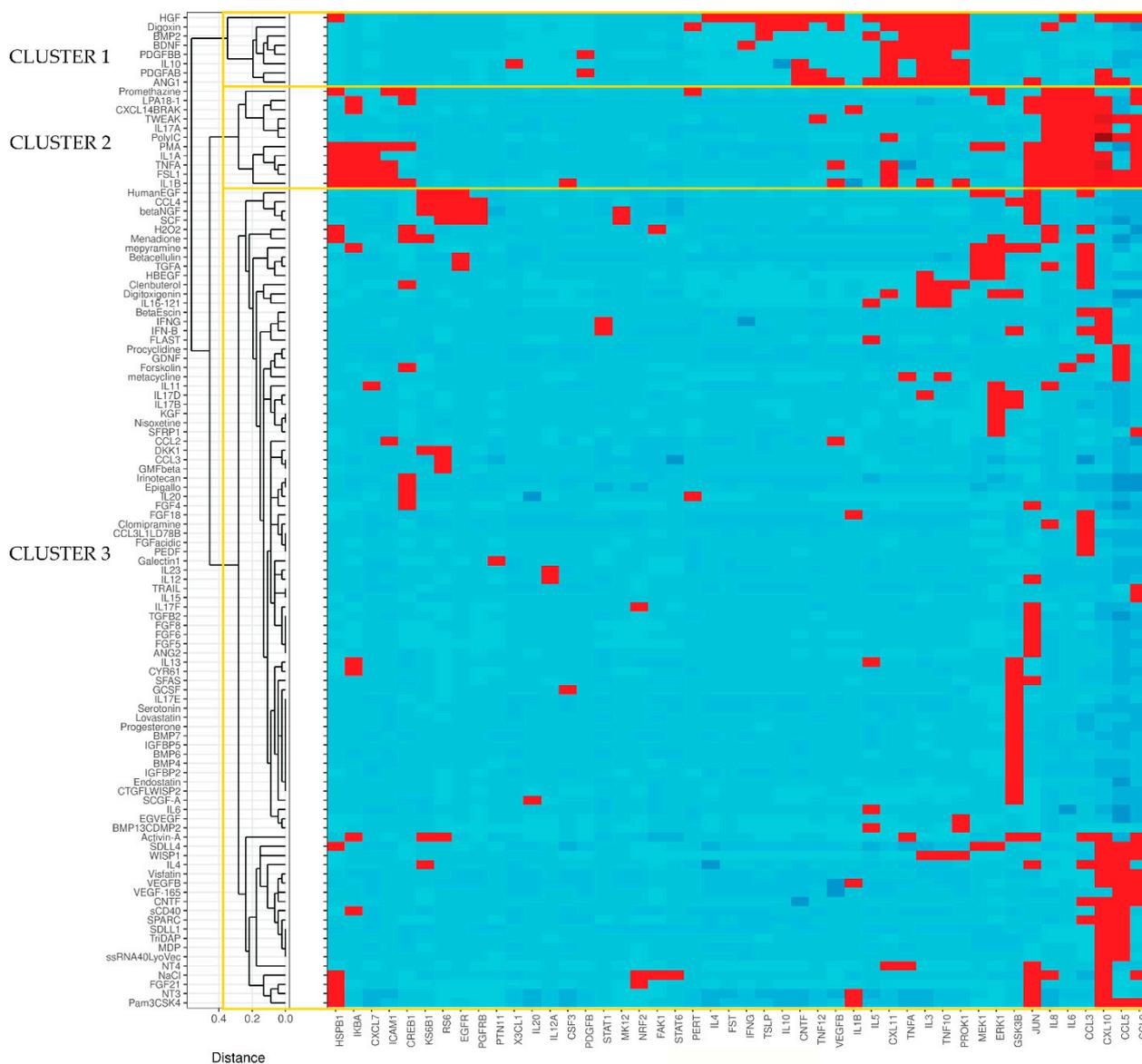


Figure 6. LPA clusters with proinflammatory stimuli. Heatmap of the active stimuli clustered in three major groups (1-3). Inactive stimuli and globally unresponsive signals were removed. Pairwise stimuli distance was calculated on binary transformed fold change values using Gower's metric prior to divisive clustering. See also Figures S2–S4.

The rest of the 108 active stimuli, out of the 176 used in our experiments, form cluster 3. Here, we are mentioning some of these stimuli that have also been linked with renal fibrosis. For example, the proinflammatory cysteine-rich protein 61 (Cyr61), which is increased in a mouse model of renal fibrosis [25], induced the phosphorylation of I κ B α and GSK3B in our experimental setup. Activin A, which activates renal interstitial fibroblasts during the fibrotic processes of the kidney [26], also triggered the phosphorylation of many intracellular components (GSK3B, KS6B1, RS6, JUN, I κ B α) and the secretion of chemokines CCL3, CXCL10, and CCL2 in HKC-8 cells. WNT1-inducible-signaling pathway protein 1 (WISP1), which has been found to regulate kidney fibrosis through TGF β [27] and whose serum levels are elevated in CKD patients [28], promoted the secretion of many factors too (CCL2, CCL5, CXCL10, TNF10, PROK1, IL-3). It would, therefore, be interesting to attempt to find a connection between some of the observed responses and renal pathology.

2.4. Pathway-Enrichment Analysis, Microarray, and Single-Cell Analysis in Relation to Cluster 2

To better characterize cluster 2, we performed pathway-enrichment analysis against both GO and KEGG libraries (see Section 4). For this purpose, we maintained the gene-coded stimuli triggering at least one response, as well as the signals responding to at least one stimulus (Supplementary Figure S2A). Top enriched terms of all GO categories and KEGG database suggest that cluster 2 is functionally related with inflammatory processes and responses to pathogen invasion (Supplementary Figure S2B,C).

Next, to prioritize some of the cluster 2 features, we re-analyzed two microarray datasets containing control as well as CKD, focal segmental glomerulosclerosis, and immunoglobulin A nephropathy samples. The differentially expressed genes identified (pathology vs. steady-state condition) include genes encoding four cluster 2-induced molecules: TNFSF12, IL-1 β , ICAM1, and CCL5 (Supplementary Figure S3). In fact, ICAM1 is upregulated upon kidney disease in all three groups of nephropathy samples, underscoring its importance in the context of kidney disease. Interestingly, according to our results, ICAM1 is activated, among others, by LPA, indicating potentially common mechanisms between LPA effects and kidney pathologies.

To examine cluster 2 features' cell specificity, we have re-analyzed a single-cell RNA-seq dataset with kidney samples from CKD and control individuals [29]. Five cytokine genes induced by cluster 2 molecules were identified as marker genes of CD10⁻ proximal tubule epithelial cells (PT-neg), *CCL2*, *CCL3*, *CCL5*, *CXCL8* (the gene encoding IL-8), and *IL-1 β* , while *VEGFB* was marking CD10⁺ proximal tubule epithelial cells (PT-pos) (Supplementary Figure S4A). Marker features did not overlap between the two PT clusters. The aforementioned results suggest that PT cells can indeed in vivo produce cluster 2-induced cytokines, several of them being responsive to LPA treatment (*CCL2*, *CCL3*, *IL-8*). Last, we performed a ligand-receptor (cell-to-cell) analysis, so as to identify the final recipient cells of these secreted cytokines. The SingleCellSignalR ligand-receptor database was enriched with specific interactions from CellTalkDB. All cells were grouped per population (epithelial, endothelial, mesenchymal, neuronal, immune) with the exception of the two PT cells' sub-clusters, which remained intact. Focus on PT-marker cytokines suggests that PT cells could establish both autocrine as well as paracrine communications with all other cell populations, such as epithelial, mesenchymal, and endothelial cells, through specific ligand-receptor interactions (Supplementary Figure S4B).

2.5. LPA Induces the Expression of the Secreted Factors *CCL2*, *CCL3*, *CXCL10*, *ICAM1*, *IL-6*, and *IL-8* at the mRNA Level in HKC-8 Cells

In order to verify the expression of the observed secreted factors upon LPA, we studied their mRNA expression. To this end, we stimulated HKC-8 cells with LPA 18:1, collected the cells, and completed the RNA extraction. Initially, we stimulated them for different time periods (0, 1, 4, 12, and 24 h). According to Figure 7A, all the analyzed secreted molecules were maximally induced at 4 h; therefore, we chose this time point for our subsequent experiments. Next, we stimulated HKC-8 cells with different concentrations of LPA (0, 2.5, 5, and 10 μ M) for 4 h and observed a dose-dependent induction of the indicated molecules with a maximum effect at 10 μ M (Figure 7B). Both the time-dependent and dose-dependent experiments verified a strong induction of *CCL2*, *CCL3*, *CXCL8*, *ICAM1*, and *IL-6* by LPA 18:1, with *CXCL8* transcription being activated over a hundred times. *CXCL10* was only marginally induced by LPA at a concentration of 2.5 μ M. Finally, we analyzed the expression of LPARs in HKC-8 cells and found that *LPAR2* is the highest expressed LPAR in these cells, followed by *LPAR6*, with no significant alterations of the LPARs pattern upon LPA stimulation (Supplementary Figure S5).

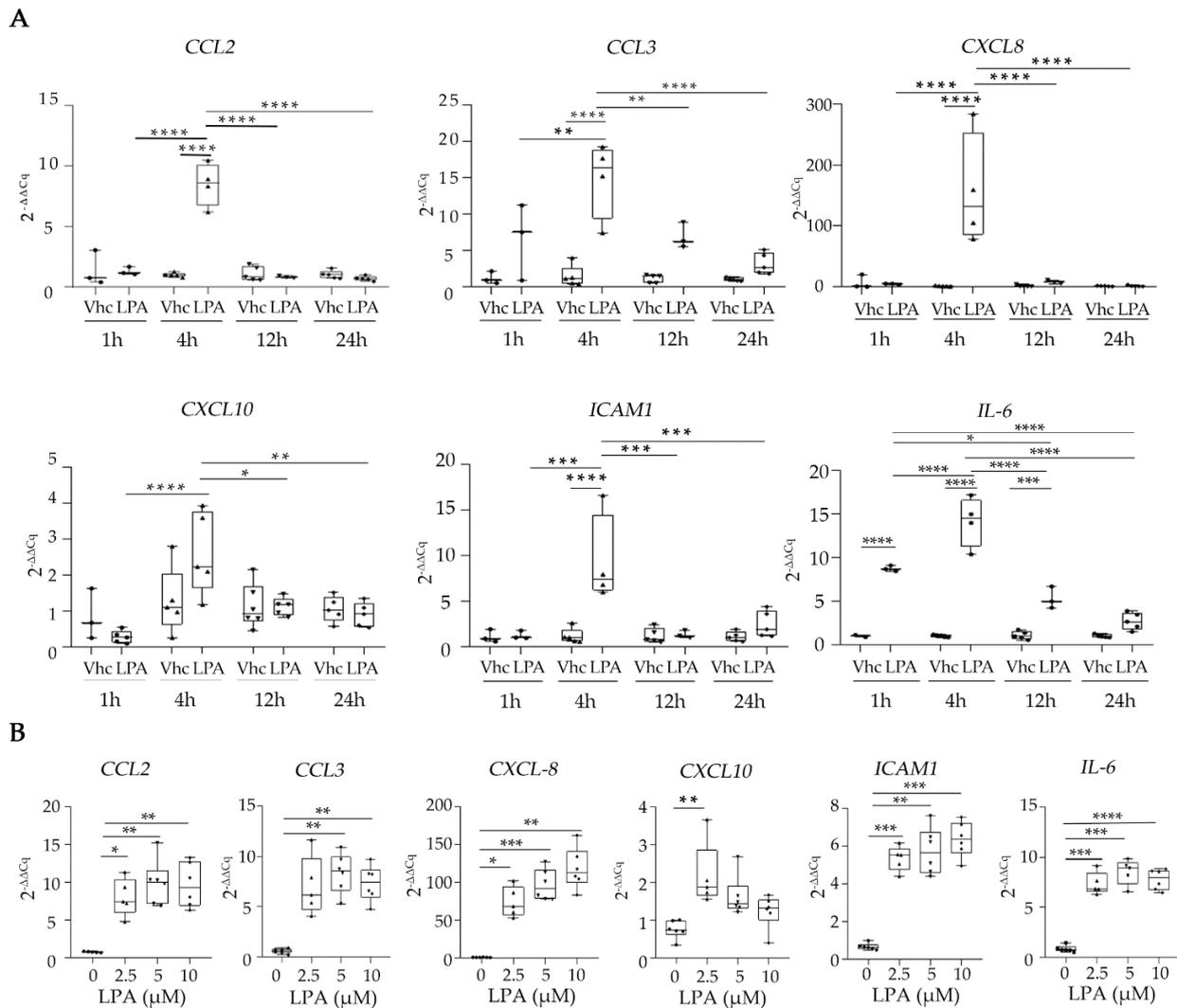


Figure 7. LPA stimulates the expression of *CCL2*, *CCL3*, *CXCL8*, *CXCL10*, *ICAM1*, and *IL6* from human kidney proximal tubular epithelial cells (HKC-8). (A,B) HKC-8 cells were incubated for 1, 4, 12, and 24 h with 10 μM of LPA (A), and with 2.5, 5, and 10 μM LPA for 4 h (B). Control cells were stimulated with the equivalent volume of chloroform (VHC). mRNA-expression levels of the indicated secreted factors were quantified with RT-qPCR. The Cq values of each gene were normalized against the Cq values of *B2M*. The results represent the findings of two (A) and three (B) separate experiments. In (A) circles, upward triangles, downward triangles and diamonds refer to 1, 4, 12 and 24 hours of incubation with LPA, respectively. In (B) circles, upward triangles, downward triangles and diamonds refer to incubation with 0, 2.5, 5 and 10 μM LPA, respectively. Statistical significance was assessed in (A) with 2-way ANOVA and Tukey's post hoc test and in (B) with Brown-Forsythe's and Welch's test or the Kruskal–Wallis test depending on the normality status of the data; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. See also Figure S5.

2.6. LPA-Induced Cellular Signaling Pathways

In order to elucidate which LPARs and signaling pathways are implicated in the LPA-mediated induction of the secreted factors, we employed several LPARs and signaling pathway inhibitors (based on the proteins that were found phosphorylated upon LPA) and checked the expression of the secreted factors, apart from the low-expressed *CXCL10*, in their presence. We identified LPAR1/3 as the receptors responsible for the LPA-induced *CCL3*, *CXCL8*, and *IL-6* transcription, while LPAR2 was responsible for the *IL-6* transcription as well (Supplementary Figure S6). In the absence of inhibitors for other LPARs, we cannot

exclude that signaling through LPAR6 or the other LPARs is participating in the regulation of the LPA-induced secreted factors. Regarding the phosphoproteins, we found that SP600125, a c-Jun N-terminal kinase (JNK) inhibitor, inhibits the expression of *CCL3* and *ICAM1* (Figure 8). Furthermore, in the presence of PD98059, a MEK/ERK inhibitor; JSH23, an NFκB inhibitor; and 666-15, a Creb inhibitor, *CCL2*, *CCL3*, *CXCL8*, and *ICAM1* expression was diminished (Figure 8). Our results suggest that the JNK/c-JUN, MEK/ERK, NFκB, and CREB pathways are implicated in the induction of some of the LPA-mediated secreted factors. In fact, *CCL2*, *CCL3*, *CXCL8*, and *ICAM1* are coregulated by these pathways simultaneously, as shown in Figure 9.

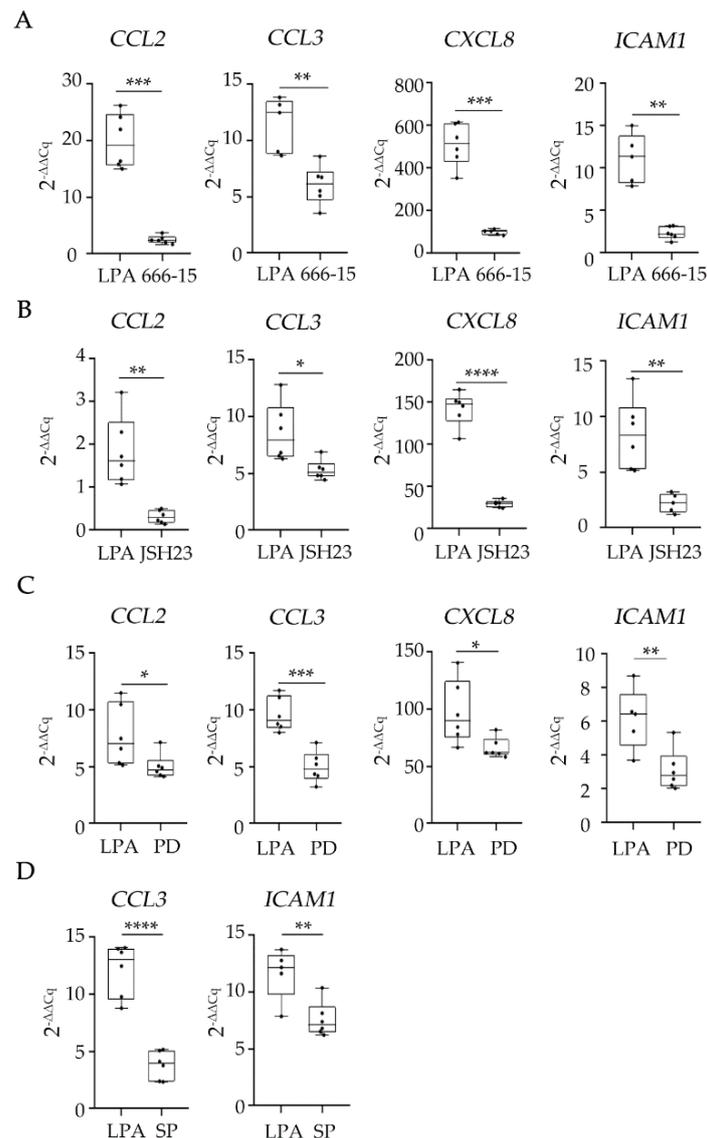


Figure 8. Pharmacologic dissection of LPA-induced cellular signaling pathways. HKC-8 cells were pretreated for 1 h with 666-15 (CREB1 inhibitor) 10 μ M in (A), JSH23 (NFκB inhibitor) 100 μ M in (B), PD98059 (MEK/ERK inhibitor) 50 μ M in (C), or SP600125 (JNK inhibitor) 50 μ M in (D) and then activated with LPA at a final concentration of 10 μ M for 4 h. mRNA-expression levels of the indicated secreted factors were quantified with RT-qPCR. The Cq values of each gene were normalized against the Cq values of *B2M*. Statistical analysis was performed with unpaired *t*-test or Welch's test in the case of normal data and with Mann–Whitney in the case of non-normal data. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. See also Figure S6.

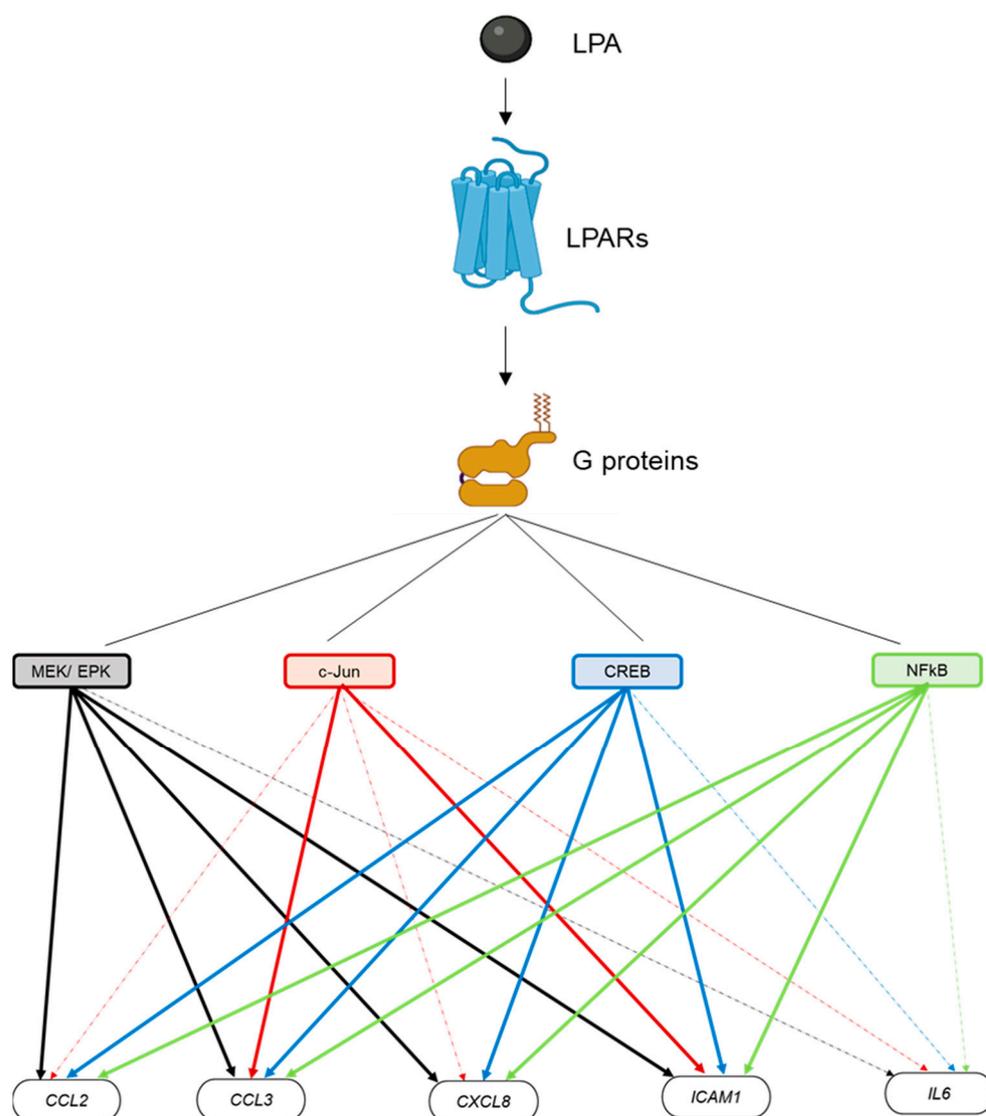


Figure 9. Graphical depiction of the LPA-induced signaling pathways in HKC-8 cells. LPA binds to the LPARs, which activate G proteins and the signal progresses to secondary signaling hubs, such as MEK/ERK or transcription factors c-JUN, CREB1, and NFκB. MEK/ERK, CREB1, and NFκB co-activate *CCL2*, *CCL3*, *CXCL8* (IL-8), and *ICAM1* expression. C-JUN activates only *CCL3* and *ICAM1* expression. Solid colored lines show connections that are derived from our results. Connections depicted with dashed lines are drawn from the literature and are not verified from our data. Created with [BioRender.com](https://www.biorender.com), accessed on 1 June 2022.

3. Discussion

In this study, we have performed custom multiplex ELISA in HKC-8 cells, a cell line of proximal tubular epithelial cells, which in vivo constitute an abundant cell population of the kidney. We have stimulated these cells with LPA and 175 immunological stimuli and monitored the phosphorylations of signaling molecules and the secretion of immune-related factors, such as cytokines, chemokines, and adhesion molecules. Out of the 176 stimuli, 108 were the active ones, with HGF evoking the most responses. Chemokine CXCL10 was the most common active responder among the secreted factors, while JUN was the signaling molecule with the most activations. Regarding LPA, which is our molecule of interest, it induced the phosphorylation of JUN, IκBα, MEK1, CREB1, and, marginally, ERK1 and the secretion of *CCL2*, *CCL3*, *CXCL10*, *ICAM1*, IL-6, and IL-8, as observed with the usage of three different LPA species. In the context of verifying the LPA-induced expression of the

six secreted factors at the mRNA level, we investigated the levels of the various LPARs in the HKC-8 cell line and found that the most prominent receptor is LPAR2, followed by LPAR6. The prevalence of LPAR2 in HKC-8 cells is corroborated by other investigations studying the presence of LPARs in mouse renal TEC cell lines [19–21]. Subsequently, we verified the LPA-induced transcription of *CCL2*, *CCL3*, *CXCL8* (the gene encoding IL-8), *ICAM1*, *IL-6*, and, marginally, *CXCL10* in these cells.

TECs are known sources of CCL chemokines [30]. These molecules have a critical role in progressive renal injury, as suggested by functional blocking studies, including treatment with neutralizing antibodies to CCL chemokines or their receptors, truncated chemokines, or small molecule-receptor antagonists [30]. They are important regulators of leukocyte recruitment during renal injury. Both CCL2, also known as monocyte chemoattractant protein 1 (MCP-1), and CCL3, also known as macrophage inflammatory protein-1 α (MIP-1 α), which are induced by LPA, form gradients that drive infiltration of monocytes/macrophages, T cells, and B cells to the sites of injury [31]. More specifically, CCL2 is released by TECs after renal injury, inducing the influx of CCL2 receptor, CCR2-positive cells such as monocytes, dendritic cells, T cells, and fibrocytes [31]. Monocytes differentiate in M1/M2 macrophages, with M1 producing proinflammatory cytokines, such as TNF α , IL-1 β , IL-6 and CCL2 and with M2 promoting wound healing and leading to TGF- β and anti-inflammatory cytokines' expression [31]. CCL2 is implicated in the pathogenesis of several diseases with a strong monocytic component. In the context of renal pathophysiology, CCL2 participates in glomerulonephritis (GN) [32,33], DN [34,35], and the CKD model of unilateral ureter obstruction (UO) [36]. Chemical or genetic ablation of CCR2 reduces renal fibrosis, TGF β production, and macrophage accumulation in several models of CKD [31]. Additionally, antagonism of CCR2 has positive effects in patients with type 2 DN [37]. LPA has been found to induce CCL2 production in mesangial cells and proximal tubular epithelial cells HK2 in vitro [38,39], corroborating the ability of LPA to promote CCL2 expression.

Regarding CCL3, this is a chemokine involved in the acute inflammatory state in the recruitment and activation of polymorphonuclear leukocytes. Elevated levels of CCL3 and its receptors CCR1 and CCR5 have been found upon UO [36]. A chemical blockade of CCR1 reduces inflammation and interstitial fibrosis in CKD murine models, such as adriamycin-induced nephropathy and UO [40,41]. To our knowledge, the induction of CCL3 by LPA has not been shown before.

CXCL10, also known as the 10-kDa interferon-inducible protein (IP-10), is a proinflammatory chemokine, as it is involved in the chemoattraction of monocytes, macrophages, T cells, and natural killer (NK) cells [42]. High CXCL10 levels have been detected in kidney biopsy specimens from patients with mesangial proliferative GN, where CXCL10 can directly contribute to mesangial cell proliferation [43]. CXCL10 levels are increased in the course of the UO model [36]. *Cxcl10*^{-/-} mice exhibit decreased proliferation with less ECM deposition and fewer cells in the glomeruli compared to wild-type mice [43]. On the contrary, recombinant murine CXCL10 reduces many indices of CKD in diabetic mice [44] and blocking CXCL10 promotes progressive renal fibrosis [45]. Therefore, more studies are needed to elucidate the role of CXCL10 in chronic renal disease.

ICAM1 or CD54 are also amplified by LPA in our experimental setup. LPA has been shown before to induce ICAM1 expression in epithelial ovarian cancer cells [46]. ICAM1 is a cell-surface glycoprotein that binds to integrins and participates in intercellular communication. Typically, it is expressed on endothelial and immune-system cells. ICAM1 is a ligand of lymphocyte-function-associated antigen-1 (LFA-1), which is a member of the integrin family found on leukocytes [47]. Via the ICAM1/LFA-1 interaction, ICAM1 stabilizes cell–cell interactions and facilitates the endothelial transmigration of leukocytes from the circulation to the sites of inflammation. ICAM1 is not detected in the TECs of healthy kidneys, whereas it is expressed in these cells upon GN [48]. Similarly, it is expressed in primary glomerulosclerosis compared to healthy renal regions, which do not express it [49]. Accordingly, the tubulointerstitial expression of ICAM1 has been

suggested as a marker of injury in IgA nephropathy [50]. The levels of serum ICAM1 are increased in diabetes and its expression has been associated with DN [51]. Based on our analysis, ICAM1 is found upregulated in microarrays from kidneys of patients with CKD, focal segmental glomerulosclerosis, and IgA nephropathy compared to controls, thus underscoring its importance in all these situations. Other cluster 2-induced molecules that are found upregulated in these microarrays are TNFSF12/TWEAK, CCL5, and IL-1 β , however, only ICAM1 is induced by LPA.

IL-6 is an interleukin that acts both as a proinflammatory cytokine and an anti-inflammatory myokine. When secreted by T cells and macrophages, it stimulates immune responses that lead to inflammation, e.g., during infection or trauma [52]. IL-6 signaling promotes T cell proliferation and apoptosis resistance; it is implicated in CD4⁺ T cell differentiation and plays a key role in the T-cell-mediated immune response, whereas it is indirectly involved in B cell-induced inflammation [53]. Therefore, IL-6 stimulates the inflammatory and auto-immune processes in many diseases. In the context of the kidney, the serum levels of IL-6 are significantly higher in CKD patients compared to healthy subjects [54] and numerous kidney resident cells, such as endothelial cells, mesangial cells, podocytes, and TECs can secrete it [53]. Several stimuli, such as glomerular injury, can induce IL-6 production from renal TECs, thus promoting a TEC–glomeruli crosstalk [53]. LPA has been shown before to induce IL-6 production in human bronchial epithelial cells [55], keratinocytes [56], and mesangial cells [38]. Exposure of mesangial cells to IL-6 and its soluble receptor (sIL-6R) together promotes the synthesis and secretion of CCL2/MCP1 and subsequently enhances monocyte recruitment [57]. Furthermore, IL-6 is implicated in fibrosis, as it can stimulate collagen I expression from TECs in vitro, while chronic administration of IL-6 enhances ischemia-reperfusion-induced fibrosis in vivo [58]. Interestingly, the blockade of IL-6 trans-signaling attenuates renal fibrosis and inflammation in the UO model of kidney fibrosis [59].

IL-8, which is the molecule most prominently induced by LPA at the mRNA level, is a key mediator associated with inflammation as it causes the activation and chemotaxis of neutrophils, leading them towards the site of inflammation [60]. LPA has been shown to induce IL-8 production before, in the bronchial epithelial cells of the lungs [55,61,62], keratinocytes [56], and epithelial ovarian cancer cells [46]. In the kidney, human renal cortical epithelial cells express IL-8 upon incubation with IL-1 β , TNF, or LPS [63]. Moreover, proximal and distal TECs are strongly positive for IL-8 in renal biopsies from patients with acute allograft rejection [63], and serum IL-8 levels are exacerbated in children with CKD [64]. In the kidneys of patients with T2 diabetes (T2D), glomerular IL-8 expression has been found to increase compared to controls [65]. Moreover, blockade of the IL-8-CXCR1/2 axis decreases diabetic-kidney-disease progression in mice [65].

Apart from CCL2, the rest of the secreted factors that we describe are shown for the first time to be LPA-induced in proximal TECs, and this may have an impact on several kidney pathologies. Given that all of the aforementioned molecules are proinflammatory, the LPA-induced secretome from TECs is characterized as proinflammatory and perhaps senescent too; five out of the six secreted factors are senescence-associated secretory phenotype proteins (SASP) [66,67]. Furthermore, the clustering of LPA with proinflammatory stimuli such as PMA, IL-1 α , IL-1 β , TNF α , IL-17 α , TWEAK/TNFSF12, and FSL1 further enhances its proinflammatory characterization. Therefore, LPA is suggested to participate in the pathology of CKD.

With regard to the phosphorylated signaling proteins, LPA promoted phosphorylation of JUN, I κ BA, CREB1, and, marginally, MEK1 and ERK. MEK1 is a signaling kinase upstream of ERK, which is an extracellular signal-regulated kinase that has been shown before to be phosphorylated in the presence of LPA in lung epithelial cells [61,68]. The three other proteins activated by LPA are transcription factors. JUN, in combination with FOS, forms the AP-1 early-response transcription factor. It is activated through double phosphorylation by the JNK pathway and is involved in cell-cycle progression and cancer. AP-1 transcribes numerous genes related to the inflammatory response, including cytokines (e.g., TNF α),

chemokines (e.g., *CCL2*), and leukocyte-adhesion molecules (e.g., *VCAM-1*) [69]. A study has illustrated before an LPA-induced phosphorylation of JUN in human bronchial epithelial cells [61]. CREB1, cAMP-responsive element-binding protein 1, is a transcription factor that binds to the cAMP-response element, a DNA nucleotide sequence present in many immune-related genes, including *IL-6* [70]. LPA has been shown to induce CREB signaling in lung epithelial cells, and the conditioned medium from these cells evokes profibrotic changes in lung fibroblasts [68]. I κ B α is the well-known inhibitor of NF κ B transcription factor, which by default sequesters NF κ B in the cytoplasm but, upon stimulation, becomes phosphorylated and allows the release of NF κ B, a central mediator of the human immune response. NF κ B, in the context of chronic inflammatory and autoimmune diseases, is activated by proinflammatory cytokines and drives proinflammatory cytokine, chemokine, and adhesion molecules' production as well [71,72]. Our Luminex results indicated activation of the NF κ B pathway in the presence of LPA. LPA treatment has been shown before to induce NF κ B in many other circumstances [73]. In the kidney, LPA increases phosphorylation of NF κ Bp65, and the LPAR1 inhibitor AM095 suppresses their activation in mesangial cells [38]. Apart from ERK, it is the first time that the aforementioned signaling molecules are found to be phosphorylated upon LPA in renal TECs.

The expression of the secreted molecules identified upon LPA stimulation could be mediated through the phosphorylated signaling hub proteins we detected. All the secreted molecules that were induced by LPA (*CCL2*, *CCL3*, *CXCL8*, *CXCL10*, *ICAM1*, *IL-6*) are among the known target genes of NF κ B [74] and, hence, their LPA induction could be mediated through it. Indeed, by using an NF κ B-specific inhibitor, we showed that LPA induces *CCL2*, *CCL3*, *CXCL-8*, and *ICAM1* genes through NF κ B. The *CXCL8* (IL-8) transcription via NF κ B has been previously shown in many cell types [62,75–79]; in fact, the IL-8 induction from NF κ B in human bronchial epithelial cells is initiated by LPA [62]. Increased expression of *CCL2*, at least in the context of pulmonary fibrosis, is induced by NF κ B (and AP-1 subunit c-JUN) [80]. According to other studies, IL-6 is one of the highest induced NF κ B-dependent cytokines in various cell types [78,81,82], however we did not observe such a regulation in the HKC-8 cell line. Concerning human primary proximal TECs, they are a potential source of IL-6, IL-8, and *CCL2* in response to various proinflammatory cytokines, such as IL-1 α and TNF α [72,83], which is validated by our results. Moreover, the IL-1 stimulation of IL-6, IL-8, and *CCL2* in primary human PTECs and HK2 TECs is NF κ B-dependent [72,84]. Additionally, NF κ B induces the expression of molecules related to leukocyte recruitment/adhesion such as *ICAM1* [85–87]. *CCL2*-mediated *ICAM1* expression in human TECs is predominantly dependent on NF κ B activation [88], while TNF α -induced activation of the *ICAM1* promoter in human endothelial cells depends on NF κ B as well [89,90]. Furthermore, oxidized LDL promotes the recruitment of NF κ B/p65 to the *ICAM1* promoter in endothelial cells [91].

We detected a regulation of *ICAM1* and *CCL3* through JNK, which is the kinase upstream of c-JUN. *ICAM1* is, indeed, regulated by c-JUN in IL-1-stimulated human primary fibroblasts [92] and TNF α -stimulated retinal-pigment epithelial cells [93]. Concerning *CCL3*, there are indications that it is expressed by palmitate and TNF α through JUN-involved signaling in THP-1 monocytic cells [94]. Even though it has been reported in the literature that *CCL2* and *CXCL8* are target genes of c-JUN, we did not verify this experimentally. Concerning CREB1, we found that it affects the transcription of *CCL2*, *CCL3*, *CXCL8*, and *ICAM1*. CREB1 is, indeed, required for the inducible transcription of *CXCL8* in monocytic cell lines [95]. Furthermore, it shows enriched binding to the promoter of *CCL2* in peripheral blood mononuclear cells [96]. LPA mediates CREB phosphorylation through mitogen- and stress-activated protein kinases, resulting in *CXCL8* and *CCL2* transcription in fibroblast-like synoviocytes [97]. CREB1 is involved in the expression of *CXCL8* and *CCL3* in neutrophils, too [98].

Apart from transcription factors, MAPK signaling is also implicated in the expression of the aforementioned secreted factors. Our results show that the expression of *CCL2*, *CCL3*, *CXCL8*, and *ICAM1* is MEK/ERK-dependent. Several publications are indicating the role

of MEK/ERK signaling in *CXCL8* expression [99–101]. The MEK/ERK pathway has also been shown to mediate CCL expression. IL-13-induced *CCL3* expression is dependent on ERK1/2 signaling in vivo [102]. LPS treatment augments *CCL3* transcription in vitro in bone-marrow-derived dendritic cells [103] and in vivo in the rat brain in a MEK/ERK-dependent fashion [102,104]. Moreover, TNF α or IL-1 β induction of *CCL3* mRNA in rat-nucleus pulposus cells is p38- and ERK-dependent [105]. Regarding IL-6, we did not observe a MEK/ERK effect on its transcription, although several pieces of data indicate such a regulation [106–108]. However, corroborating our results, LPA-mediated IL-6 expression is not affected by a MEK inhibitor in microglia [109]. MEK proteins also seem to control *CXCL10* expression [42,109], although we did not test this.

However, several genes encoding cytokines are simultaneously regulated by multiple signaling pathways and transcription factors. We found that *CXCL8* is regulated by MEK/ERK, NF κ B, and CREB. Indeed, CREB and NF κ B are among the transcription factors that are cooperatively activated for *CXCL8* transcription in human bronchial epithelial cells [77]. MEK/ERK and transcription factors AP-1 and NF κ B are all involved in *CXCL8* upregulation by IL-1 β in gastric-carcinoma cells [99] and by CD40 in human fetal microglia [100]. Moreover, the *P. aeruginosa*-dependent transcription of *CXCL8* in human bronchial epithelial cells is mediated by ERK signaling and a multitude of transcription factors, such as NF κ B, AP-1, and CREB [77]. ICAM1 expression is controlled both by NF κ B and c-JUN, upon PMA or TNF α , in endothelial cells [110]. As other studies suggest, ICAM1 expression is also regulated by both NF κ B and CREB [111]. Both NF κ B and ERK1/2 mediate *CCL3*/*MIP-1a* expression in the brain [104], something that we also verify in the HKC-8 cell line along with co-regulation by the JNK/JUN and CREB pathways. *CCL2* and *CCL3*, among other chemokines, are induced by H₂O₂ through ERK and the nuclear translocation of NF κ B, AP-1 and CREB in macrophages [112]. We also show that *CCL3* is co-regulated by MEK/ERK, JNK/JUN, NF κ B, and CREB in the HKC-8 cells, whereas *CCL2* is activated by MEK/ERK, NF κ B, and CREB.

Employing divisive clustering for all the tested stimuli, LPA congregates with IL-1 α , IL-1 β , IL-17 α , TNF α , TNFSF12/TWEAK, and *CXCL14*/*BRAK*, which are endogenous stimuli in the human body, and promethazine, PMA, FSL1, and PolyIC, which are exogenous stimuli. All the aforementioned endogenous stimuli are implicated in CKD. High levels of IL-1 α are detected in renal TECs in biopsies from DN patients, while in vitro IL-1 α provokes the deposition of ECM proteins [113]. IL-1 β mRNA is also detected in biopsies of DN patients [114]. IL-1 β contributes to systemic inflammation and the progression of modeled CKD, either type 2 diabetes-induced or adenine diet-induced, as shown by studies utilizing monoclonal anti-IL-1 β in mice [114,115]. IL-1 β is also produced by several cell types during IgA nephropathy and promotes inflammation and disease progression [116]. Besides, activation of the inflammasome in immune cells during kidney injury causes the secretion of IL-1 α and IL-1 β , which then promote cytokine and chemokine release through the IL-1 receptor (IL-1R), resulting in the further recruitment of immune cells [117]; thus, these two cytokines are important in the inflammatory component of kidney disease [118]. It is, therefore, well expected that the deletion of type I IL-1R ameliorates the early renal fibrosis induced by ureter obstruction in mice [119].

IL-17 α levels are increased in the kidneys of diabetic mice compared to control kidneys, and treatment with anti-IL-17 α antibody ameliorates renal dysfunction and disease [120]. Additionally, IL-17 α positive cells have been detected in renal biopsies of hypertensive nephroangiosclerosis and kidneys of experimental hypertensive mice, while mice infused with IL-17 α show higher inflammatory cell infiltration in the kidneys, with a simultaneous elevated *CCL2* and *CCL5* gene expression [121]. Moreover, it is postulated that IL-17 α promotes the AKI-to-CKD transition [122]. However, studies claiming an antifibrotic role of IL-17 α also exist [123,124].

TNFSF12/TWEAK is another molecule that clusters with LPA. TNFSF12/TWEAK was found to induce the secretion of *CCL2*, *CCL3*, *CCL5*, IL-6, IL-8, and *CXCL10*. Three of these molecules (*CCL2*, *CCL5*, and IL-6) have been previously shown to increase upon

TNFSF12/TWEAK treatment in renal tubular cells [125], thus corroborating our results. TNFSF12/TWEAK is also implicated in renal injury [126–128]. Although its expression is rather low in normal kidneys, it becomes significant during tissue damage in diverse forms of AKI and CKD [129]. Increased protein expression of TNFSF12/TWEAK is detected in the renal cortex of patients with lupus nephritis (LN) and treatment of mesangial cells with TNFSF12/TWEAK promotes macrophage chemotaxis, probably through the chemotactic factors that TNFSF12/TWEAK induces [130]. Indeed, TNFSF12/TWEAK promotes the NF κ B-mediated expression of proinflammatory cytokines, such as CCL2 and CCL5, in human glomerular mesangial cells [127]. Inhibiting TNFSF12/TWEAK in vivo reduces tubular chemokine expression and macrophage infiltration [125]. Blocking or deleting TNFSF12/TWEAK or its receptor induces a drop in inflammation and an improvement of renal function in several experimental models of renal disease [129,131]. On the contrary, in vivo TNFSF12/TWEAK administration leads to NF κ B activation in the whole kidney and expression of chemokines from tubular cells [132]. Moreover, the development of anti-TWEAK therapies against inflammatory diseases such as RA is in progress [129].

TNF α is also implicated in renal function, as it directly affects the hemodynamic and excretory function of the kidney [133]. It is a potent proinflammatory cytokine, which, however, also has an immunosuppressive effect. In healthy kidneys, the levels of TNF α are very low, whereas they increase in many kidney diseases upon leukocyte infiltration, as activated monocytes and macrophages are its primary source [134]. In terms of its expression, TNF α is not only expressed by the infiltrating macrophages but also by resident kidney cells such as mesangial, podocytes, and TECs [134]. TNF α can regulate proliferation and apoptosis in renal cells, but it can elicit a local proinflammatory cytokine cascade, too. In UUO, renal TNF production is increased after ureter obstruction and is implicated in tubular-cell apoptosis and interstitial fibrosis [134]. Additionally, TNF α is increased in patients with acute allograft rejection and chronic allograft nephropathy [135]. Intriguingly, in RA, TNF α has been found to induce ATX expression from synovial fibroblasts (SFs), while it induces SF activation and effector functions in synergy with LPA [136], thus proposing a possible synergism of TNF α with LPA in CKD, too.

CXCL14/BRAK promotes chemotaxis of immature dendritic cells, neutrophils, monocytes, activated human NK cells, and others [137]. CXCL14 has been detected in kidney specimens, however, it has not been extensively investigated, except for a study where CXCL14 overexpression mitigates sepsis-induced AKI, probably through the regulation of the M1/M2 macrophage ratio and the downregulation of cytokine production [138].

As most stimuli in the LPA-including cluster 2 promote CKD, we assume that LPA has a negative impact on CKD as well. Indeed, the role of LPA in CKD has been established by several studies on the mouse models of various renal pathologies. LPA is significantly increased in the urine of mice subjected to the CKD model of nephrectomy [139]. Upon UUO, LPA production from kidney explants is enhanced and LPAR1 is found to be upregulated (although LPAR2 and LPAR6 are, by default, the highest expressed LPA receptors in the kidney) [20,21]. Importantly, genetic deletion or pharmacological suppression of LPAR1 reduces tubulointerstitial fibrotic and inflammatory markers in mice subjected to UUO [20,21,140]. LPAR1 ablation decreases the number of proliferating fibroblasts and accumulating myofibroblasts induced by UUO [21]. In vitro, LPA is shown to induce CTGF expression in mouse primary proximal TECs through LPAR1 and LPAR2; CTGF then stimulates fibroblast proliferation and their differentiation to myofibroblasts, thus promoting epithelial–fibroblast communication [21].

DN is another manifestation of CKD, where LPA and the ATX/LPA/LPAR axis are implicated. LPA and LPC are significantly increased in the renal glomeruli of eNOS^(-/-) db/db mice, a robust model of DN [141]. In the same model, LPAR1, LPAR3, and ATX-expression levels are upregulated upon disease; administration of the LPAR1/LPAR3 antagonist BMS002 ameliorates glomerular filtration and renal fibrosis, while it reduces macrophage infiltration and podocyte loss [18]. In a similar db/db model of T2D, both ATX and LPAR1 are overexpressed in the kidney cortex compared to control mice [142]. Simulta-

neously, another inhibitor of LPAR1/LPAR3 ameliorates albuminuria and glomerulosclerosis, the main pathological feature of type 2 DN [142]. Similarly, an LPAR1 inhibitor, AM095, inhibits the expression of proinflammatory cytokines and fibrotic factors in the kidney, reduces glomerular matrix expansion, and improves kidney function in a streptozotocin-induced type 1 diabetic model [38]. In a mesangial cell line, LPA significantly increases the expression of proinflammatory cytokines TNF α , IL-1 β , IL-6, and CCL2/MCP-1 and promotes phosphorylation of NF κ B and JNK [38], while at the same time it induces the profibrotic factors TGF β 1 and fibronectin in a glycogen synthase kinase (GSK)3B and sterol regulatory element-binding protein (SREBP1)-mediated fashion [142]. In our study, we have gone one step further, showing that LPA promotes the expression of proinflammatory molecules in TECs as well.

Beyond cluster 2, the stimulus with the most responders was HGF, which belongs to cluster 1. Intriguingly, HGF seems to suppress chronic renal failure, and administration of HGF improves renal fibrosis [143]. Another molecule of interest is one of the most common secreted factors identified in our experiments, TNFSF10/TRAIL. Experimental and clinical studies have illustrated that TNFSF10/TRAIL is up-regulated in different kidney diseases, both in DN and in non-diabetic conditions such as LN, rejected kidney transplant, AKI, and others [129]. The TNFSF10/TRAIL receptor, TRAIL-R2, has been pointed out as the protein most strongly associated with the decline of kidney function [129].

Finally, apart from the previously mentioned instances, a number of our results can be verified by the existing data in the literature. According to a study, the addition of LPA in HKC-8 cells induces ERK1/2 phosphorylation in these cells [144], corroborating our results where LPA 18:1 induced phosphorylation of MEK, of the kinase upstream of ERK, and of ERK, albeit not statistically significantly. Other studies also support our results, as HKC-8 cells have been found to respond to BMP-7 by reversing TGF β 1-induced EMT [145,146]. In our assay, HKC-8 cells also respond to BMP-7 by inducing phosphorylation of GSK3B. In an in vitro model studying kidney fibrosis, it has been found that upon injury with cisplatin, HKC-8 cells secrete CCL5 and IL-6 [147]. Our Luminex data indeed show that HKC-8 cells are able to secrete these cytokines, as both CCL5 and IL-6 were induced upon a series of stimuli. In another setup, EGF activated EGFR, p38 MAPK, NF κ Bp65, and STAT3, leading to inducible nitric oxide synthase expression in HKC-8 cells [148]. In our experiment, human EGF drove many protein phosphorylations, including EGFR, JUN, MEK1, ERK1, and others. However, it did not activate STAT3 or MK12 (p38 γ).

Conclusively, in this report, we have identified the responses of human renal proximal TECs to a series of 176 immunological stimuli. The subsequent quantification of the levels of 27 intracellular phosphoproteins and 32 extracellular molecules with multiplex ELISA reveals a large amount of information concerning the signaling and physiology of renal proximal tubular epithelial cells and their possible interaction with resident stromal cells. Among the 176 stimuli, LPA stands out as a proinflammatory stimulus promoting the phosphorylation of important signaling hubs and the secretion of factors of clinical relevance concerning CKD. Finally, our results offer some mechanistic insight into the contribution of LPA to kidney-related chronic inflammation and further pinpoint the ATX/LPA axis as crucial in the development of renal pathology and as a possible therapeutic target.

Limitations of the Study

Among the limitations of the study is the fact that the main multiple ELISA experiment with the 176 stimuli was performed in single wells, instead of triplicates, however, this was a high-throughput experiment that would be difficult to have been performed in triplicates. Moreover, the data of the LPA response and the response on some other stimuli have been verified with further experiments done in triplicate wells. Another limitation is the fact that the inhibitors used against LPARs or signaling molecules may have been added in excess, despite our efforts to adhere to concentrations used in several other publications. Finally, we realize that our experiments were done in vitro, and, therefore, in an artificial environment deprived of the extracellular matrix and neighboring cells, which poses another limitation.

4. Materials and Methods

4.1. Cell Culture and Cell Stimulation with 176 Immune Molecules

HKC-8 cells, provided by Roche, were grown in DMEM:F12, glutamine 2 mM, insulin-transferrin-selenium (ITS) supplement 1×, FBS 2.5%, penicillin 100 u/mL, streptomycin 100 µg/mL, and amphotericin B 2.5 µg/mL. They were seeded at a cell density of 24,000 cells/well in 96-well plates and left to attach overnight. Starvation followed next day with DMEM:F12, glutamine 2 mM, ITS supplement 1×, BSA 0.2%, penicillin 100 u/mL, and streptomycin 100 µg/mL for 3 h. The medium was replaced (again with starvation medium) and the addition of multiple stimuli followed, as shown in the Supplementary Materials. Two experiments took place. The first experiment was designed to measure the intracellular phosphorylation events; hence, we used cell lysates. The second experiment was designed to measure secreted factors such as cytokines; therefore, we used cell supernatants. The phosphorylation events were tested at two time points, meaning that this experiment involved two sub-experiments.

The stimuli were added at the same concentrations between the two experiments but at different volumes. In total, 5 µL of diluted stimuli were added in 20 µL of medium in each well of the phosphoprotein plates, and 20 µL of stimuli were added in 80 µL of medium in each well of the secreted factors' plates. In total, 175 stimuli were added apart from LPA and the controls. For each experiment (phosphoproteins/secreted factors), the 175 stimuli and LPA 18:1 were added in single wells, while medium (as control) was added in 5 wells separated in the 2 plates. For the repetition of LPA stimulation, three LPA species were used, each of them added in triplicate wells and chloroform (LPA's solvent) again in triplicate wells. Before the addition, LPA and chloroform were heated so as to be easily diluted in the cell medium.

Between the phosphoprotein and secreted-factors experiments, the incubation time with the stimuli differed: 5 min for one set of plates for the phosphoprotein experiment, 25 min for another set of plates for the phosphoprotein experiment, and 24 h for the secreted factors experiment. The different incubation times reflect the different times necessary for phosphorylation events and expression (transcription and translation) to take place.

4.2. Multiplex ELISA

4.2.1. Phosphoprotein Experiment

After the 5 or 25-min incubation period, each plate was placed on an ice pack in order to stop the reactions in all wells simultaneously, and cells were lysed by adding 40 µL of lysis buffer mix in each well. The lysis buffer mix contained a ProtATonce custom-made lysis buffer and phosphatase inhibitors, a protease inhibitors cocktail, and extra phenylmethylsulfonyl fluoride (PMSF). Lysis took place by shaking the plate at 4 °C for 20 min while keeping it continuously on the ice pack. The plates with the lysates were stored at −20 °C, covered with aluminum plate sealers. Prior to the Luminex assay, the lysates were thawed and sonicated (4 times, 10 s each) and the plates were spun down at 2700 g for 20 min. The top 50 µL of the samples were transferred into flat-bottom 96-well plates containing 50 µL of bead mix per well, pre-washed with Assay Buffer (PBS with 1% BSA, pH 7.4). The bead mix contained magnetic beads internally dyed with precise proportions of red and infrared fluorophores. The differing proportions of the red and infrared fluorophores result in unique spectral signature microspheres. Each unique microsphere-bead was conjugated with a distinct mAb against a phosphoprotein, thus allowing simultaneous recognition of 27 phosphoproteins in one sample.

Each plate, covered with a sealer, was shaken at maximum speed for 90 min and then placed on a magnetic separator that keeps the magnetic beads down and allows discarding of the supernatant. The bead-sample mix was washed twice with the assay buffer in this manner and the biotinylated second/detection antibody was added in all wells. Incubation with the second antibody lasted 90 min, with the sealed plates shaking at maximum speed. Extra assay buffer was added and the beads were washed twice in order to remove the excess antibody. A Streptavidin and R-Phycoerythrin Conjugate mix diluted in assay buffer

was used at 5 µg/mL, and a volume of 50 µL was added per well. Following a 15 min shaking incubation, the supernatant was discarded, the beads were washed with assay buffer, and, finally, 130 µL of assay buffer was added per well prior to the measurement.

4.2.2. Secreted-Factors Experiment

After the 24-h incubation, the plates were spun down (200 g for 5 min) and the supernatants were transferred into new plates, covered with an aluminum plate sealer and stored at −20 °C. On the day of the Luminex assay, 50 µL of the samples were transferred into flat-bottom 96-well plates, containing the bead mix conjugated with mAbs against 32 secreted factors. The same process followed as described for the phosphoproteins.

All measurements were taken in a Flexmap 3D of Luminex corporation. Significant effort was devoted to maximizing the number of measurements that could be obtained from each sample of cells: a 96-well plate assayed for 27 phosphoproteins yielded 2592 measurements, and a plate assayed for 32 secreted factors yielded 3072 measurements.

4.3. Bioinformatic Analysis

4.3.1. Luminex-Data Preprocessing

Independent analysis was done for the phosphoprotein experiment and the secreted factors experiment. In both experiments, the control wells, treated with plain medium, were distributed between the different plates and the median of the multiple control wells was calculated. The response to the plain medium (control wells) was considered as the unstimulated state, whereas the response to a stimulus was considered as the stimulated state. For each phosphoprotein or secreted factor induced by a certain stimulus, the measurement at the stimulated state (usually a single measurement or the median value in case of replicates, e.g., the triplicate LPA wells for each LPA species) was divided with the measurement at the unstimulated state (median of the control wells). The emerging ratio corresponds to the fold-change in the response of each measured component (phosphoprotein/ secreted factor) to the specific stimulus, compared to the response to plain medium. The distribution of the fold changes seen in the responses is depicted in Supplementary Figure S1A.

In order to call a signal (the ratio of stimulated to unstimulated state) active or not, we used a threshold of fold change at 1.5. The choice of threshold was made through a sensitivity analysis on the effect of the threshold on the signals dataset. In particular, the percentage of activations was recorded at several thresholds. At low thresholds, slight changes in the threshold greatly affected the number of activations, whereas, at higher thresholds, the dataset was rather insensitive to threshold changes. Therefore, we decided to set the threshold at 1.5, where the number of activations is rather stable, but there are also enough activations for the subsequent analysis (Supplementary Figure S1B).

All the responses were processed using the open-access MATLAB-based software DataRail (<http://code.google.com/p/sbpipeline/wiki/DataRail> (version v1.3, accessed on 1 June 2014)).

4.3.2. Clustering

Phosphoprotein and secreted factors fold changes were combined and stimuli not causing any signal activation were removed, along with globally non-responsive signals. Gower's metric was used to define pairwise-stimuli distances post fold change to binary values' transformation (1: activated; 0: non-activated). Divisive clustering was performed based on the calculated distance matrix. Gower's distance was calculated using the proxy R package, while divisive clustering was performed using cluster R package functions.

4.3.3. Pathway Analysis

Cluster 2 stimuli causing at least one activation event and signals responding at least once were concatenated. Elements not being coded by a gene were filtered out, with the exception of LPA 18:1, which was replaced by *ENPP2*. ClusterProfiler R package [149] was used for over-representation analysis of both GO terms and KEGG pathways using

default parameters. An FDR-corrected p -value threshold of at least 0.05 was applied to define significantly enriched terms.

4.3.4. Microarray-Data Re-Analysis

Raw microarray data were fetched from GEO series GSE66494 (Agilent 4x44K G4112F) and GSE104066 (Affymetrix HuGene-2_1-st) using the GEOquery R package [150]. Agilent microarrays were background corrected using the normexp method with a 50 offset, as suggested by limma package authors [151], and then quantile normalized between arrays. Affymetrix data were background corrected and RMA normalized using the oligo R package [152]. Both datasets were quality-controlled post normalization, using arrayQualityMetrics R package [153] and Principal Component Analysis (PCA). One (GSM1623315) and two (GSM3904846, GSM2788881) outlier samples were filtered from GSE66494 and GSE104066, respectively. Control probes and probes matching to either no or multiple HGNC gene symbols were not considered for downstream analysis, along with those having intensity values close to the background. Agilent probes with a high cross-linking potential were not maintained either. Probe-intensity values were summarized at the gene level, and weighted mean value was calculated in one:many gene:probe relationships. Differential-expression analysis was performed using an empirical Bayes statistic as implemented in limma package. Absolute fold change ≥ 1.2 and FDR corrected p -value < 0.05 were set as differential expression thresholds.

4.3.5. Single-Cell Data Analysis

Single-cell data were downloaded from zenodo (<https://doi.org/10.5281/zenodo.4059315> (accessed on 15 November 2021)). Pre-processed CD10⁺ and CD10⁻ objects were log-normalized and integrated using Seurat package v4.0.5 using a pre-computed AnchorSet based on 2000 features [154]. Biological annotation of the original publication was maintained. Marker genes and differential expression analysis was performed using FindMarkers function under default parameters on the RNA assay of the integrated object. Absolute fold change ≥ 1.2 and adjusted p -value < 0.05 were set as significance thresholds.

Cell-to-cell analysis was performed by SingleCellSignalR package v1.4.0 [155]. Cells were grouped per cellular population, except for proximal tubule positive (PT-pos; CD10⁺) and proximal tubule negative (PT-neg; CD10⁻) cells. SingleCellSignalR default-ligand-receptor database was extended to include some interactions of interest, as described in CellTalkDB v0.0.1 [156]. Circos plots were visualized using circlize package v0.4.13 [157].

4.4. RNA Analysis

4.4.1. LPA Stimulation—RNA Isolation—Reverse Transcription—RT-qPCR

HKC-8 cells were seeded in 6-well plates at a density equal to 300,000 cells/well with a subsequent overnight starvation (medium with 0.2% BSA and no FBS). Cells were stimulated with 2.5, 5, or 10 μ M LPA 18:1 (Avanti Lipids, Sigma-Aldrich, Merck, St Louis, MO, USA), dissolved in starvation medium. As a control, plain starvation medium or starvation medium with the LPA's solvent, chloroform, was used. Cells were incubated at 37 °C and 5% CO₂ for 1, 4, 12, or 24 h prior to RNA extraction. RNA extraction was done with TRI Reagent (TR118, MRC, Cincinnati, OH, USA) in accordance with the instructions of the manufacturer, with a slight modification at the RNA-precipitation stage, where glycogen was added. The RNA concentration and purity were determined with NanoDrop[®] ND-1000 UV-Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), calculating the optical density ratio at wavelengths of 260/280 nm and 260/230 nm. Samples were placed at -80 °C until further use. First-strand cDNA was generated with the Moloney murine leukemia virus reverse transcriptase (28025-013, Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) using 2 μ g of RNA and according to the reagent's protocol. Real-time PCR was performed on a BioRad CFX96 Touch[™] Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA), using SYBR Select Master Mix (4472913, Thermo Fisher Scientific, Waltham, MA, USA), 25 ng of each cDNA per reaction, and primers that

are listed in Table S3. The thermal-cycling conditions for 40-cycles amplification were at 95 °C for 10 s and 58, 59, or 60 °C for 45 s. Normalization of the Ct values was done against the reference gene *B2M*. The relative quantification of the target-gene expression was done using the Livak ($2^{-\Delta\Delta Cq}$) method and presented as fold change of each normalized target gene in the LPA-treated samples relative to control samples. The statistical analysis between groups of the dose-response study was performed using Brown–Forsythe’s and Welch’s ANOVA tests or the Kruskal–Wallis test in case of non-normal distribution. Two-way-ANOVA was performed for the time-course experiments. Finally, in the case of LPARs expression, the $2^{-\Delta Cq}$ formula was used in order to compare the levels of the different LPARs. Statistical analysis was done with GraphPad.

4.4.2. Experiments with LPAR or Phosphoprotein Inhibitors

HKC-8 cells were seeded and starved as above and pretreated with inhibitors for one hour. LPAR1/3 inhibitor (Ki16425, Cat. no: HY-13285, MedChemExpress, Monmouth Junction, NJ, USA) was added at 10 μ M; LPAR2 inhibitor (H2L5186303, Cat. no: 10-1452, Focus Biomolecules, Plymouth Meeting, PA, USA) was added at 10 μ M; CREB inhibitor (666-15, Cat. no: A616443, Toronto Research Chemicals, North York, ON, Canada) was added at 10 μ M; JNK inhibitor (SP600125, Cat. no: 420119, Calbiochem, San Diego, CA, USA) was added at 50 μ M; NF κ B inhibitor (JSH-23, Cat. no: HY-13982, MedChemExpress, Monmouth Junction, NJ, USA) was added at 100 μ M; and MEK/ERK inhibitor (PD98059, Cat. no: 513000, Calbiochem, San Diego, CA, USA) was added at 50 μ M. As controls, some wells were treated with plain medium or with an equivalent volume of DMSO, the solvent of the inhibitors. After 1 h, and without removing the inhibitor or DMSO, LPA was added to the experimental wells to a final concentration of 10 μ M for 4 h. RNA isolation, reverse transcription, and RT-qPCR were performed as above. The statistical analysis was done with unpaired *t*-test or Welch’s *t*-test, depending on the equality of standard deviation between the different groups, or Mann–Whitney in the case of non-normal data. Statistical analysis was done with GraphPad.

4.5. Image Creation

Images of Figures 1 and 9 were created with [BioRender.com](https://www.bio-render.com/), with agreement numbers OH23ZPFMT5 and CX23ZPEU47, respectively. [BioRender.com](https://www.bio-render.com/) was accessed on 1 June 2022.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms23137452/s1>.

Author Contributions: Conceptualization, V.A., L.G.A. and C.U.; methodology, L.G.A., D.F., T.S., G.A. and C.M.; software, T.S. and D.F.; validation, V.P., G.A. and C.M.; formal analysis, T.S., D.F., G.A. and C.M.; investigation, V.P., G.A. and C.M.; resources, V.A., L.G.A. and C.U.; data curation, T.S. and D.F.; writing—original draft preparation, C.M., D.F. and G.A.; writing—review and editing, V.A.; visualization, T.S. and D.F.; supervision, V.A. and L.G.A.; project administration, V.A. and L.G.A.; funding acquisition, V.A. and C.U. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been co-financed by the European Union and Greek National Funds through the Operational Programme Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH—CREATE—INNOVATE (project code: T1EDK-00949). The project has been supported from an unrestricted research grant from F. Hoffmann-La Roche AG.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: C.U. is an employee of F. Hoffmann-La Roche AG.

Abbreviations

ANG1: angiopoietin; ARPCs, adult renal progenitor cells; ATX, autotaxin; BDNF, brain-derived neurotrophic factor; BMP2, bone morphogenetic protein 2; CKD, chronic kidney disease; CNTF, Ciliary Neurotrophic Factor; CREB1, cAMP responsive element binding protein 1; CTGF, connective tissue growth factor; DN, diabetic nephropathy; ECM, extracellular matrix; EMT, epithelial to mesenchymal transition; FC, fold change; GFR, glomerular-filtration rate; GN, glomerulonephritis; GSK, glycogen synthase kinase; HGF, hepatocyte growth factor; ICAM1, Intercellular Adhesion Molecule 1; IL, interleukin; IP-10, 10-kDa interferon-inducible protein; ITS, insulin-transferrin-selenium; JNK, c-Jun N-terminal kinase; LFA1, lymphocyte-function-associated antigen-1; LN, lupus nephritis; LPA, lysophosphatidic acid; LPAR, LPA receptor; MCP1, monocyte chemoattractant protein-1; MIP-1 α , macrophage inflammatory protein-1 α ; NK, natural killer; PDGFB, platelet-derived growth factor; PMA, phorbol 12-myristate 13-acetate; PROK1, Prokineticin 1; PT, proximal tubule; RA, rheumatoid arthritis; RANTES, regulated on activation, normal T cell expressed and secreted; SASP, senescence-associated secretory phenotype proteins; SFs, synovial fibroblasts; SREBP1, sterol regulatory element-binding protein; T2D, type 2 diabetes; TECs, tubular epithelial cells; TGF β , transforming growth factor beta; TLR, toll-like receptor; TNF α , tumor necrosis factor alpha; TSLP, thymic stromal lymphopoietin; UUO, unilateral ureter obstruction; WISP1, WNT1-inducible-signaling pathway protein 1.

References

1. Lv, J.C.; Zhang, L.X. Prevalence and Disease Burden of Chronic Kidney Disease. *Adv. Exp. Med. Biol.* **2019**, *1165*, 3–15. [[CrossRef](#)] [[PubMed](#)]
2. Romagnani, P.; Remuzzi, G.; Glassock, R.; Levin, A.; Jager, K.J.; Tonelli, M.; Massy, Z.; Wanner, C.; Anders, H.J. Chronic kidney disease. *Nat. Rev. Dis. Primers* **2017**, *3*, 17088. [[CrossRef](#)]
3. Ruiz-Ortega, M.; Rayego-Mateos, S.; Lamas, S.; Ortiz, A.; Rodrigues-Diez, R.R. Targeting the progression of chronic kidney disease. *Nat. Rev. Nephrol.* **2020**, *16*, 269–288. [[CrossRef](#)]
4. Takaori, K.; Nakamura, J.; Yamamoto, S.; Nakata, H.; Sato, Y.; Takase, M.; Nameta, M.; Yamamoto, T.; Economides, A.N.; Kohno, K.; et al. Severity and Frequency of Proximal Tubule Injury Determines Renal Prognosis. *J. Am. Soc. Nephrol.* **2016**, *27*, 2393–2406. [[CrossRef](#)] [[PubMed](#)]
5. Ferenbach, D.A.; Bonventre, J.V. Mechanisms of maladaptive repair after AKI leading to accelerated kidney ageing and CKD. *Nat. Rev. Nephrol.* **2015**, *11*, 264–276. [[CrossRef](#)] [[PubMed](#)]
6. Zeisberg, M.; Neilson, E.G. Mechanisms of tubulointerstitial fibrosis. *J. Am. Soc. Nephrol.* **2010**, *21*, 1819–1834. [[CrossRef](#)]
7. Yang, L.; Besschetnova, T.Y.; Brooks, C.R.; Shah, J.V.; Bonventre, J.V. Epithelial cell cycle arrest in G2/M mediates kidney fibrosis after injury. *Nat. Med.* **2010**, *16*, 535–543. [[CrossRef](#)]
8. Qi, R.; Yang, C. Renal tubular epithelial cells: The neglected mediator of tubulointerstitial fibrosis after injury. *Cell Death Dis.* **2018**, *9*, 1126. [[CrossRef](#)]
9. Breda, P.C.; Wiech, T.; Meyer-Schwesinger, C.; Grahammer, F.; Huber, T.; Panzer, U.; Tiegs, G.; Neumann, K. Renal proximal tubular epithelial cells exert immunomodulatory function by driving inflammatory CD4⁺ T cell responses. *Am. J. Physiol.* **2019**, *317*, F77–F89. [[CrossRef](#)]
10. Gewin, L.; Zent, R.; Pozzi, A. Progression of chronic kidney disease: Too much cellular talk causes damage. *Kidney Int.* **2017**, *91*, 552–560. [[CrossRef](#)]
11. Rapa, S.F.; Di Iorio, B.R.; Campiglia, P.; Heidland, A.; Marzocco, S. Inflammation and Oxidative Stress in Chronic Kidney Disease-Potential Therapeutic Role of Minerals, Vitamins and Plant-Derived Metabolites. *Int. J. Mol. Sci.* **2019**, *21*, 263. [[CrossRef](#)] [[PubMed](#)]
12. Cachofeiro, V.; Goicochea, M.; de Vinuesa, S.G.; Oubina, P.; Lahera, V.; Luno, J. Oxidative stress and inflammation, a link between chronic kidney disease and cardiovascular disease. *Kidney Int.* **2008**, *74*, S4–S9. [[CrossRef](#)] [[PubMed](#)]
13. Gupta, J.; Mitra, N.; Kanetsky, P.A.; Devaney, J.; Wing, M.R.; Reilly, M.; Shah, V.O.; Balakrishnan, V.S.; Guzman, N.J.; Girndt, M.; et al. Association between albuminuria, kidney function, and inflammatory biomarker profile in CKD in CRIC. *Clin. J. Am. Soc. Nephrol.* **2012**, *7*, 1938–1946. [[CrossRef](#)] [[PubMed](#)]
14. Magkrioti, C.; Galaris, A.; Kanellopoulou, P.; Stylianaki, E.A.; Kaffe, E.; Aidinis, V. Autotaxin and chronic inflammatory diseases. *J. Autoimmun.* **2019**, *104*, 102327. [[CrossRef](#)] [[PubMed](#)]
15. Sasagawa, T.; Suzuki, K.; Shiota, T.; Kondo, T.; Okita, M. The significance of plasma lysophospholipids in patients with renal failure on hemodialysis. *J. Nutr. Sci. Vitaminol.* **1998**, *44*, 809–818. [[CrossRef](#)]
16. Zhang, Z.H.; Chen, H.; Vaziri, N.D.; Mao, J.R.; Zhang, L.; Bai, X.; Zhao, Y.Y. Metabolomic Signatures of Chronic Kidney Disease of Diverse Etiologies in the Rats and Humans. *J. Proteome Res.* **2016**, *15*, 3802–3812. [[CrossRef](#)]

17. Saulnier-Blache, J.S.; Feigerlova, E.; Halimi, J.M.; Gourdy, P.; Roussel, R.; Guerci, B.; Dupuy, A.; Bertrand-Michel, J.; Bascands, J.L.; Hadjadj, S.; et al. Urinary lysophospholipids are increased in diabetic patients with nephropathy. *J. Diabetes Its Complicat.* **2017**, *31*, 1103–1108. [[CrossRef](#)] [[PubMed](#)]
18. Zhang, M.Z.; Wang, X.; Yang, H.; Fogo, A.B.; Murphy, B.J.; Kaltenbach, R.; Cheng, P.; Zinker, B.; Harris, R.C. Lysophosphatidic Acid Receptor Antagonism Protects against Diabetic Nephropathy in a Type 2 Diabetic Model. *J. Am. Soc. Nephrol.* **2017**, *28*, 3300–3311. [[CrossRef](#)]
19. Geng, H.; Lan, R.; Singha, P.K.; Gilchrist, A.; Weinreb, P.H.; Violette, S.M.; Weinberg, J.M.; Saikumar, P.; Venkatachalam, M.A. Lysophosphatidic acid increases proximal tubule cell secretion of profibrotic cytokines PDGF-B and CTGF through LPA2- and Galphaq-mediated Rho and alphavbeta6 integrin-dependent activation of TGF-beta. *Am. J. Pathol.* **2012**, *181*, 1236–1249. [[CrossRef](#)]
20. Pradere, J.P.; Klein, J.; Gres, S.; Guigne, C.; Neau, E.; Valet, P.; Calise, D.; Chun, J.; Bascands, J.L.; Saulnier-Blache, J.S.; et al. LPA1 receptor activation promotes renal interstitial fibrosis. *J. Am. Soc. Nephrol.* **2007**, *18*, 3110–3118. [[CrossRef](#)]
21. Sakai, N.; Chun, J.; Duffield, J.S.; Lagares, D.; Wada, T.; Luster, A.D.; Tager, A.M. Lysophosphatidic acid signaling through its receptor initiates profibrotic epithelial cell fibroblast communication mediated by epithelial cell derived connective tissue growth factor. *Kidney Int.* **2017**, *91*, 628–641. [[CrossRef](#)] [[PubMed](#)]
22. Racusen, L.C.; Monteil, C.; Sgrignoli, A.; Lucskay, M.; Marouillat, S.; Rhim, J.G.; Morin, J.P. Cell lines with extended in vitro growth potential from human renal proximal tubule: Characterization, response to inducers, and comparison with established cell lines. *J. Lab. Clin. Med.* **1997**, *129*, 318–329. [[CrossRef](#)]
23. He, F.F.; Zhang, D.; Chen, Q.; Zhao, Y.; Wu, L.; Li, Z.Q.; Zhang, C.; Jiang, Z.H.; Wang, Y.M. Angiopoietin-Tie signaling in kidney diseases: An updated review. *FEBS Lett.* **2019**, *593*, 2706–2715. [[CrossRef](#)] [[PubMed](#)]
24. Simone, S.; Cosola, C.; Loverre, A.; Cariello, M.; Sallustio, F.; Rascio, F.; Gesualdo, L.; Schena, F.P.; Grandaliano, G.; Pertosa, G. BMP-2 induces a profibrotic phenotype in adult renal progenitor cells through Nox4 activation. *Am. J. Physiol.* **2012**, *303*, F23–F34. [[CrossRef](#)] [[PubMed](#)]
25. Lai, C.F.; Chen, Y.M.; Chiang, W.C.; Lin, S.L.; Kuo, M.L.; Tsai, T.J. Cysteine-rich protein 61 plays a proinflammatory role in obstructive kidney fibrosis. *PLoS ONE* **2013**, *8*, e56481. [[CrossRef](#)]
26. Yamashita, S.; Maeshima, A.; Kojima, I.; Nojima, Y. Activin A is a potent activator of renal interstitial fibroblasts. *J. Am. Soc. Nephrol.* **2004**, *15*, 91–101. [[CrossRef](#)]
27. Wang, B.; Ding, X.; Ding, C.; Tesch, G.; Zheng, J.; Tian, P.; Ricardo, S.; Shen, H.H.; Xue, W. WNT1-inducible-signaling pathway protein 1 regulates the development of kidney fibrosis through the TGF-beta1 pathway. *FASEB J.* **2020**, *34*, 14507–14520. [[CrossRef](#)]
28. Zhong, X.; Tu, Y.J.; Li, Y.; Zhang, P.; Wang, W.; Chen, S.S.; Li, L.; Chung, A.C.; Lan, H.Y.; Chen, H.Y.; et al. Serum levels of WNT1-inducible signaling pathway protein-1 (WISP-1): A noninvasive biomarker of renal fibrosis in subjects with chronic kidney disease. *Am. J. Transl. Res.* **2017**, *9*, 2920–2932.
29. Kuppe, C.; Ibrahim, M.M.; Kranz, J.; Zhang, X.; Ziegler, S.; Perales-Patón, J.; Jansen, J.; Reimer, K.C.; Smith, J.R.; Dobie, R.; et al. Decoding myofibroblast origins in human kidney fibrosis. *Nature* **2021**, *589*, 281–286. [[CrossRef](#)]
30. Chung, A.C.; Lan, H.Y. Chemokines in renal injury. *J. Am. Soc. Nephrol.* **2011**, *22*, 802–809. [[CrossRef](#)]
31. Lv, W.; Booz, G.W.; Wang, Y.; Fan, F.; Roman, R.J. Inflammation and renal fibrosis: Recent developments on key signaling molecules as potential therapeutic targets. *Eur. J. Pharmacol.* **2018**, *820*, 65–76. [[CrossRef](#)] [[PubMed](#)]
32. Lloyd, C.M.; Minto, A.W.; Dorf, M.E.; Proudfoot, A.; Wells, T.N.; Inlamat, D.J.; Gutierrez-Ramos, J.C. RANTES and monocyte chemoattractant protein-1 (MCP-1) play an important role in the inflammatory phase of crescentic nephritis, but only MCP-1 is involved in crescent formation and interstitial fibrosis. *J. Exp. Med.* **1997**, *185*, 1371–1380. [[CrossRef](#)] [[PubMed](#)]
33. Wada, T.; Yokoyama, H.; Furuichi, K.; Kobayashi, K.I.; Harada, K.; Naruto, M.; Su, S.B.; Akiyama, M.; Mukaida, N.; Matsushima, K. Intervention of crescentic glomerulonephritis by antibodies to monocyte chemotactic and activating factor (MCAF/MCP-1). *FASEB J.* **1996**, *10*, 1418–1425. [[CrossRef](#)] [[PubMed](#)]
34. Lee, S.Y.; Kim, S.I.; Choi, M.E. Therapeutic targets for treating fibrotic kidney diseases. *Transl. Res.* **2015**, *165*, 512–530. [[CrossRef](#)] [[PubMed](#)]
35. Chow, F.Y.; Nikolic-Paterson, D.J.; Ozols, E.; Atkins, R.C.; Rollin, B.J.; Tesch, G.H. Monocyte chemoattractant protein-1 promotes the development of diabetic renal injury in streptozotocin-treated mice. *Kidney Int.* **2006**, *69*, 73–80. [[CrossRef](#)]
36. Vielhauer, V.; Anders, H.J.; Mack, M.; Cihak, J.; Strutz, F.; Stangassinger, M.; Luckow, B.; Grone, H.J.; Schlondorff, D. Obstructive nephropathy in the mouse: Progressive fibrosis correlates with tubulointerstitial chemokine expression and accumulation of CC chemokine receptor 2- and 5-positive leukocytes. *J. Am. Soc. Nephrol.* **2001**, *12*, 1173–1187. [[CrossRef](#)]
37. De Zeeuw, D.; Bekker, P.; Henkel, E.; Hasslacher, C.; Gouni-Berthold, I.; Mehling, H.; Potarca, A.; Tesar, V.; Heerspink, H.J.; Schall, T.J.; et al. The effect of CCR2 inhibitor CCX140-B on residual albuminuria in patients with type 2 diabetes and nephropathy: A randomised trial. *Lancet. Diabetes Endocrinol.* **2015**, *3*, 687–696. [[CrossRef](#)]
38. Lee, J.H.; Sarker, M.K.; Choi, H.; Shin, D.; Kim, D.; Jun, H.S. Lysophosphatidic acid receptor 1 inhibitor, AM095, attenuates diabetic nephropathy in mice by downregulation of TLR4/NF-kappaB signaling and NADPH oxidase. *Biochim. Biophys. Acta Mol. Basis Dis.* **2019**, *1865*, 1332–1340. [[CrossRef](#)]
39. Zheng, Z.; Li, C.; Shao, G.; Li, J.; Xu, K.; Zhao, Z.; Zhang, Z.; Liu, J.; Wu, H. Hippo-YAP/MCP-1 mediated tubular maladaptive repair promote inflammation in renal failed recovery after ischemic AKI. *Cell Death Dis.* **2021**, *12*, 754. [[CrossRef](#)]

40. Anders, H.J.; Vielhauer, V.; Frink, M.; Linde, Y.; Cohen, C.D.; Blattner, S.M.; Kretzler, M.; Strutz, F.; Mack, M.; Grone, H.J.; et al. A chemokine receptor CCR-1 antagonist reduces renal fibrosis after unilateral ureter ligation. *J. Clin. Investig.* **2002**, *109*, 251–259. [[CrossRef](#)]
41. Vielhauer, V.; Berning, E.; Eis, V.; Kretzler, M.; Segerer, S.; Strutz, F.; Horuk, R.; Grone, H.J.; Schlondorff, D.; Anders, H.J. CCR1 blockade reduces interstitial inflammation and fibrosis in mice with glomerulosclerosis and nephrotic syndrome. *Kidney Int.* **2004**, *66*, 2264–2278. [[CrossRef](#)] [[PubMed](#)]
42. Vazirinejad, R.; Ahmadi, Z.; Kazemi Arababadi, M.; Hassanshahi, G.; Kennedy, D. The biological functions, structure and sources of CXCL10 and its outstanding part in the pathophysiology of multiple sclerosis. *Neuroimmunomodulation* **2014**, *21*, 322–330. [[CrossRef](#)] [[PubMed](#)]
43. Gao, J.; Wu, L.; Wang, S.; Chen, X. Role of Chemokine (C-X-C Motif) Ligand 10 (CXCL10) in Renal Diseases. *Mediat. Inflamm.* **2020**, *2020*, 6194864. [[CrossRef](#)] [[PubMed](#)]
44. Zhang, Y.; Thai, K.; Kepecs, D.M.; Winer, D.; Gilbert, R.E. Reversing CXCL10 Deficiency Ameliorates Kidney Disease in Diabetic Mice. *Am. J. Pathol.* **2018**, *188*, 2763–2773. [[CrossRef](#)] [[PubMed](#)]
45. Nakaya, I.; Wada, T.; Furuichi, K.; Sakai, N.; Kitagawa, K.; Yokoyama, H.; Ishida, Y.; Kondo, T.; Sugaya, T.; Kawachi, H.; et al. Blockade of IP-10/CXCR3 promotes progressive renal fibrosis. *Nephron. Exp. Nephrol.* **2007**, *107*, e12–e21. [[CrossRef](#)]
46. Dutta, S.; Wang, F.Q.; Wu, H.S.; Mukherjee, T.J.; Fishman, D.A. The NF-kappaB pathway mediates lysophosphatidic acid (LPA)-induced VEGF signaling and cell invasion in epithelial ovarian cancer (EOC). *Gynecol. Oncol.* **2011**, *123*, 129–137. [[CrossRef](#)]
47. Staunton, D.E.; Marlin, S.D.; Stratowa, C.; Dustin, M.L.; Springer, T.A. Primary structure of ICAM-1 demonstrates interaction between members of the immunoglobulin and integrin supergene families. *Cell* **1988**, *52*, 925–933. [[CrossRef](#)]
48. Muller, G.A.; Markovic-Lipkovski, J.; Muller, C.A. Intercellular adhesion molecule-1 expression in human kidneys with glomerulonephritis. *Clin. Nephrol.* **1991**, *36*, 203–208.
49. Dal Canton, A.; Fuiano, G.; Sepe, V.; Caglioti, A.; Ferrone, S. Mesangial expression of intercellular adhesion molecule-1 in primary glomerulosclerosis. *Kidney Int.* **1992**, *41*, 951–955. [[CrossRef](#)]
50. Arrizabalaga, P.; Sole, M.; Abellana, R.; de las Cuevas, X.; Soler, J.; Pascual, J.; Ascaso, C. Tubular and interstitial expression of ICAM-1 as a marker of renal injury in IgA nephropathy. *Am. J. Nephrol.* **2003**, *23*, 121–128. [[CrossRef](#)]
51. Gu, H.F.; Ma, J.; Gu, K.T.; Brismar, K. Association of intercellular adhesion molecule 1 (ICAM1) with diabetes and diabetic nephropathy. *Front. Endocrinol.* **2012**, *3*, 179. [[CrossRef](#)] [[PubMed](#)]
52. Pedersen, B.K.; Febbraio, M.A. Muscle as an endocrine organ: Focus on muscle-derived interleukin-6. *Physiol. Rev.* **2008**, *88*, 1379–1406. [[CrossRef](#)] [[PubMed](#)]
53. Su, H.; Lei, C.T.; Zhang, C. Interleukin-6 Signaling Pathway and Its Role in Kidney Disease: An Update. *Front. Immunol.* **2017**, *8*, 405. [[CrossRef](#)] [[PubMed](#)]
54. Oh, D.J.; Kim, H.R.; Lee, M.K.; Woo, Y.S. Profile of human beta-defensins 1,2 and proinflammatory cytokines (TNF-alpha, IL-6) in patients with chronic kidney disease. *Kidney Blood Press. Res.* **2013**, *37*, 602–610. [[CrossRef](#)]
55. Barekzi, E.; Roman, J.; Hise, K.; Georas, S.; Steinke, J.W. Lysophosphatidic acid stimulates inflammatory cascade in airway epithelial cells. *Prostagland. Leukot. Essent. Fat. Acids* **2006**, *74*, 357–363. [[CrossRef](#)]
56. Thorlakson, H.H.; Engen, S.A.; Schreurs, O.; Schenck, K.; Blix, I.J.S. Lysophosphatidic acid induces expression of genes in human oral keratinocytes involved in wound healing. *Arch. Oral Biol.* **2017**, *80*, 153–159. [[CrossRef](#)]
57. Coletta, I.; Soldo, L.; Polentarutti, N.; Mancini, F.; Guglielmotti, A.; Pinza, M.; Mantovani, A.; Milanese, C. Selective induction of MCP-1 in human mesangial cells by the IL-6/sIL-6R complex. *Exp. Nephrol.* **2000**, *8*, 37–43. [[CrossRef](#)]
58. Ranganathan, P.; Jayakumar, C.; Ramesh, G. Proximal tubule-specific overexpression of netrin-1 suppresses acute kidney injury-induced interstitial fibrosis and glomerulosclerosis through suppression of IL-6/STAT3 signaling. *Am. J. Physiol.* **2013**, *304*, F1054–F1065. [[CrossRef](#)]
59. Chen, W.; Yuan, H.; Cao, W.; Wang, T.; Chen, W.; Yu, H.; Fu, Y.; Jiang, B.; Zhou, H.; Guo, H.; et al. Blocking interleukin-6 trans-signaling protects against renal fibrosis by suppressing STAT3 activation. *Theranostics* **2019**, *9*, 3980–3991. [[CrossRef](#)]
60. Bernhard, S.; Hug, S.; Stratmann, A.E.P.; Erber, M.; Vidoni, L.; Knapp, C.L.; Thomass, B.D.; Fauler, M.; Nilsson, B.; Nilsson Ekdahl, K.; et al. Interleukin 8 Elicits Rapid Physiological Changes in Neutrophils That Are Altered by Inflammatory Conditions. *J. Innate Immun.* **2021**, *13*, 225–241. [[CrossRef](#)]
61. Saatian, B.; Zhao, Y.; He, D.; Georas, S.N.; Watkins, T.; Spannhake, E.W.; Natarajan, V. Transcriptional regulation of lysophosphatidic acid-induced interleukin-8 expression and secretion by p38 MAPK and JNK in human bronchial epithelial cells. *Biochem. J.* **2006**, *393*, 657–668. [[CrossRef](#)] [[PubMed](#)]
62. Cummings, R.; Zhao, Y.; Jacoby, D.; Spannhake, E.W.; Ohba, M.; Garcia, J.G.; Watkins, T.; He, D.; Saatian, B.; Natarajan, V. Protein kinase Cdelta mediates lysophosphatidic acid-induced NF-kappaB activation and interleukin-8 secretion in human bronchial epithelial cells. *J. Biol. Chem.* **2004**, *279*, 41085–41094. [[CrossRef](#)] [[PubMed](#)]
63. Schmodder, R.L.; Strieter, R.M.; Wiggins, R.C.; Chensue, S.W.; Kunkel, S.L. In vitro and in vivo interleukin-8 production in human renal cortical epithelia. *Kidney Int.* **1992**, *41*, 191–198. [[CrossRef](#)] [[PubMed](#)]
64. Tuncay, S.C.; Dogan, E.; Hakverdi, G.; Tutar, Z.U.; Mir, S. Interleukin-8 is increased in chronic kidney disease in children, but not related to cardiovascular disease. *Braz. J. Nephrol.* **2021**, *43*, 359–364. [[CrossRef](#)]
65. Loretelli, C.; Rocchio, F.; D’Addio, F.; Ben Nasr, M.; Castillo-Leon, E.; Dellepiane, S.; Vergani, A.; Abdelsalam, A.; Assi, E.; Maestroni, A.; et al. The IL-8-CXCR1/2 axis contributes to diabetic kidney disease. *Metabolism* **2021**, *121*, 154804. [[CrossRef](#)]

66. Malaquin, N.; Martinez, A.; Rodier, F. Keeping the senescence secretome under control: Molecular reins on the senescence-associated secretory phenotype. *Exp. Gerontol.* **2016**, *82*, 39–49. [[CrossRef](#)]
67. Schafer, M.J.; Zhang, X.; Kumar, A.; Atkinson, E.J.; Zhu, Y.; Jachim, S.; Mazula, D.L.; Brown, A.K.; Berning, M.; Aversa, Z.; et al. The senescence-associated secretome as an indicator of age and medical risk. *JCI Insight* **2020**, *5*, e133668. [[CrossRef](#)]
68. Nathan, S.; Zhang, H.; Andreoli, M.; Leopold, P.L.; Crystal, R.G. CREB-dependent LPA-induced signaling initiates a pro-fibrotic feedback loop between small airway basal cells and fibroblasts. *Respir. Res.* **2021**, *22*, 97. [[CrossRef](#)]
69. Grynberg, K.; Ma, F.Y.; Nikolic-Paterson, D.J. The JNK Signaling Pathway in Renal Fibrosis. *Front. Physiol.* **2017**, *8*, 829. [[CrossRef](#)]
70. Wen, A.Y.; Sakamoto, K.M.; Miller, L.S. The role of the transcription factor CREB in immune function. *J. Immunol.* **2010**, *185*, 6413–6419. [[CrossRef](#)]
71. Hayden, M.S.; Ghosh, S. NF-kappaB, the first quarter-century: Remarkable progress and outstanding questions. *Genes Dev.* **2012**, *26*, 203–234. [[CrossRef](#)] [[PubMed](#)]
72. De Haij, S.; Woltman, A.M.; Bakker, A.C.; Daha, M.R.; van Kooten, C. Production of inflammatory mediators by renal epithelial cells is insensitive to glucocorticoids. *Br. J. Pharmacol.* **2002**, *137*, 197–204. [[CrossRef](#)] [[PubMed](#)]
73. Sun, W.; Yang, J. Molecular basis of lysophosphatidic acid-induced NF-kappaB activation. *Cell. Signal.* **2010**, *22*, 1799–1803. [[CrossRef](#)]
74. Pahl, H.L. Activators and target genes of Rel/NF-kappaB transcription factors. *Oncogene* **1999**, *18*, 6853–6866. [[CrossRef](#)] [[PubMed](#)]
75. Simone, R.E.; Russo, M.; Catalano, A.; Monego, G.; Froehlich, K.; Boehm, V.; Palozza, P. Lycopene inhibits NF-kB-mediated IL-8 expression and changes redox and PPARgamma signalling in cigarette smoke-stimulated macrophages. *PLoS ONE* **2011**, *6*, e19652. [[CrossRef](#)] [[PubMed](#)]
76. Al-Mohanna, F.; Saleh, S.; Parhar, R.S.; Collison, K. IL-12-dependent nuclear factor-kappaB activation leads to de novo synthesis and release of IL-8 and TNF-alpha in human neutrophils. *J. Leukoc. Biol.* **2002**, *72*, 995–1002. [[PubMed](#)]
77. Bezzerri, V.; Borgatti, M.; Finotti, A.; Tamanini, A.; Gambari, R.; Cabrini, G. Mapping the transcriptional machinery of the IL-8 gene in human bronchial epithelial cells. *J. Immunol.* **2011**, *187*, 6069–6081. [[CrossRef](#)]
78. Georganas, C.; Liu, H.; Perlman, H.; Hoffmann, A.; Thimmapaya, B.; Pope, R.M. Regulation of IL-6 and IL-8 expression in rheumatoid arthritis synovial fibroblasts: The dominant role for NF-kappa B but not C/EBP beta or c-Jun. *J. Immunol.* **2000**, *165*, 7199–7206. [[CrossRef](#)]
79. Elliott, C.L.; Allport, V.C.; Loudon, J.A.; Wu, G.D.; Bennett, P.R. Nuclear factor-kappa B is essential for up-regulation of interleukin-8 expression in human amnion and cervical epithelial cells. *Mol. Hum. Reprod.* **2001**, *7*, 787–790. [[CrossRef](#)]
80. Deng, X.; Xu, M.; Yuan, C.; Yin, L.; Chen, X.; Zhou, X.; Li, G.; Fu, Y.; Feghali-Bostwick, C.A.; Pang, L. Transcriptional regulation of increased CCL2 expression in pulmonary fibrosis involves nuclear factor-kappaB and activator protein-1. *Int. J. Biochem. Cell Biol.* **2013**, *45*, 1366–1376. [[CrossRef](#)]
81. Liu, T.; Zhang, L.; Joo, D.; Sun, S.C. NF-kappaB signaling in inflammation. *Signal Transduct. Target. Ther.* **2017**, *2*, 17023. [[CrossRef](#)] [[PubMed](#)]
82. Vanden Berghe, W.; Plaisance, S.; Boone, E.; De Bosscher, K.; Schmitz, M.L.; Fiers, W.; Haegeman, G. p38 and extracellular signal-regulated kinase mitogen-activated protein kinase pathways are required for nuclear factor-kappaB p65 transactivation mediated by tumor necrosis factor. *J. Biol. Chem.* **1998**, *273*, 3285–3290. [[CrossRef](#)] [[PubMed](#)]
83. Gerritsma, J.S.; Hiemstra, P.S.; Gerritsen, A.F.; Prodjosudjadi, W.; Verweij, C.L.; Van Es, L.A.; Daha, M.R. Regulation and production of IL-8 by human proximal tubular epithelial cells in vitro. *Clin. Exp. Immunol.* **1996**, *103*, 289–294. [[CrossRef](#)] [[PubMed](#)]
84. De Haij, S.; Bakker, A.C.; van der Geest, R.N.; Haegeman, G.; Vanden Berghe, W.; Aarbiou, J.; Daha, M.R.; van Kooten, C. NF-kappaB mediated IL-6 production by renal epithelial cells is regulated by c-jun NH2-terminal kinase. *J. Am. Soc. Nephrol.* **2005**, *16*, 1603–1611. [[CrossRef](#)] [[PubMed](#)]
85. Brasier, A.R. The nuclear factor-kappaB-interleukin-6 signalling pathway mediating vascular inflammation. *Cardiovasc. Res.* **2010**, *86*, 211–218. [[CrossRef](#)]
86. Xue, J.; Thippgowda, P.B.; Hu, G.; Bachmaier, K.; Christman, J.W.; Malik, A.B.; Tiruppathi, C. NF-kappaB regulates thrombin-induced ICAM-1 gene expression in cooperation with NFAT by binding to the intronic NF-kappaB site in the ICAM-1 gene. *Physiol. Genom.* **2009**, *38*, 42–53. [[CrossRef](#)]
87. Melotti, P.; Nicolis, E.; Tamanini, A.; Rolfini, R.; Pavirani, A.; Cabrini, G. Activation of NF-kB mediates ICAM-1 induction in respiratory cells exposed to an adenovirus-derived vector. *Gene Ther.* **2001**, *8*, 1436–1442. [[CrossRef](#)]
88. Viedt, C.; Orth, S.R. Monocyte chemoattractant protein-1 (MCP-1) in the kidney: Does it more than simply attract monocytes? *Nephrol. Dial. Transplant.* **2002**, *17*, 2043–2047. [[CrossRef](#)]
89. Ledebur, H.C.; Parks, T.P. Transcriptional regulation of the intercellular adhesion molecule-1 gene by inflammatory cytokines in human endothelial cells. Essential roles of a variant NF-kappa B site and p65 homodimers. *J. Biol. Chem.* **1995**, *270*, 933–943. [[CrossRef](#)]
90. Zhou, Z.; Connell, M.C.; MacEwan, D.J. TNFR1-induced NF-kappaB, but not ERK, p38MAPK or JNK activation, mediates TNF-induced ICAM-1 and VCAM-1 expression on endothelial cells. *Cell. Signal.* **2007**, *19*, 1238–1248. [[CrossRef](#)]
91. Li, N.; Liu, H.; Xue, Y.; Chen, J.; Kong, X.; Zhang, Y. Upregulation of Neogenin-1 by a CREB1-BAF47 Complex in Vascular Endothelial Cells is Implicated in Atherogenesis. *Front. Cell Dev. Biol.* **2022**, *10*, 803029. [[CrossRef](#)] [[PubMed](#)]

92. Holzberg, D.; Knight, C.G.; Dittrich-Breiholz, O.; Schneider, H.; Dorrie, A.; Hoffmann, E.; Resch, K.; Kracht, M. Disruption of the c-JUN-JNK complex by a cell-permeable peptide containing the c-JUN delta domain induces apoptosis and affects a distinct set of interleukin-1-induced inflammatory genes. *J. Biol. Chem.* **2003**, *278*, 40213–40223. [[CrossRef](#)] [[PubMed](#)]
93. Lee, I.T.; Liu, S.W.; Chi, P.L.; Lin, C.C.; Hsiao, L.D.; Yang, C.M. TNF-alpha mediates PKCdelta/JNK1/2/c-Jun-dependent monocyte adhesion via ICAM-1 induction in human retinal pigment epithelial cells. *PLoS ONE* **2015**, *10*, e0117911. [[CrossRef](#)]
94. Sindhu, S.; Akhter, N.; Wilson, A.; Thomas, R.; Arefanian, H.; Al Madhoun, A.; Al-Mulla, F.; Ahmad, R. MIP-1alpha Expression Induced by Co-Stimulation of Human Monocytic Cells with Palmitate and TNF-alpha Involves the TLR4-IRF3 Pathway and Is Amplified by Oxidative Stress. *Cells* **2020**, *9*, 1799. [[CrossRef](#)] [[PubMed](#)]
95. Koga, Y.; Tsurumaki, H.; Aoki-Saito, H.; Sato, M.; Yatomi, M.; Takehara, K.; Hisada, T. Roles of Cyclic AMP Response Element Binding Activation in the ERK1/2 and p38 MAPK Signalling Pathway in Central Nervous System, Cardiovascular System, Osteoclast Differentiation and Mucin and Cytokine Production. *Int. J. Mol. Sci.* **2019**, *20*, 1346. [[CrossRef](#)]
96. Tomalka, J.A.; Pelletier, A.N.; Fourati, S.; Latif, M.B.; Sharma, A.; Furr, K.; Carlson, K.; Lifton, M.; Gonzalez, A.; Wilkinson, P.; et al. The transcription factor CREB1 is a mechanistic driver of immunogenicity and reduced HIV-1 acquisition following ALVAC vaccination. *Nat. Immunol.* **2021**, *22*, 1294–1305. [[CrossRef](#)]
97. Zhao, C.; Hui, W.; Fernandes, M.J.; Poubelle, P.E.; Bourgoin, S.G. Lysophosphatidic acid-induced IL-8 secretion involves MSK1 and MSK2 mediated activation of CREB1 in human fibroblast-like synoviocytes. *Biochem. Pharmacol.* **2014**, *90*, 62–72. [[CrossRef](#)]
98. Mayer, T.Z.; Simard, F.A.; Cloutier, A.; Vardhan, H.; Dubois, C.M.; McDonald, P.P. The p38-MSK1 signaling cascade influences cytokine production through CREB and C/EBP factors in human neutrophils. *J. Immunol.* **2013**, *191*, 4299–4307. [[CrossRef](#)]
99. Hwang, Y.S.; Jeong, M.; Park, J.S.; Kim, M.H.; Lee, D.B.; Shin, B.A.; Mukaida, N.; Ellis, L.M.; Kim, H.R.; Ahn, B.W.; et al. Interleukin-1beta stimulates IL-8 expression through MAP kinase and ROS signaling in human gastric carcinoma cells. *Oncogene* **2004**, *23*, 6603–6611. [[CrossRef](#)]
100. D'Aversa, T.G.; Eugenin, E.A.; Berman, J.W. CD40-CD40 ligand interactions in human microglia induce CXCL8 (interleukin-8) secretion by a mechanism dependent on activation of ERK1/2 and nuclear translocation of nuclear factor-kappaB (NFkappaB) and activator protein-1 (AP-1). *J. Neurosci. Res.* **2008**, *86*, 630–639. [[CrossRef](#)]
101. Namba, S.; Nakano, R.; Kitanaka, T.; Kitanaka, N.; Nakayama, T.; Sugiya, H. ERK2 and JNK1 contribute to TNF-alpha-induced IL-8 expression in synovial fibroblasts. *PLoS ONE* **2017**, *12*, e0182923. [[CrossRef](#)] [[PubMed](#)]
102. Lee, P.J.; Zhang, X.; Shan, P.; Ma, B.; Lee, C.G.; Homer, R.J.; Zhu, Z.; Rincon, M.; Mossman, B.T.; Elias, J.A. ERK1/2 mitogen-activated protein kinase selectively mediates IL-13-induced lung inflammation and remodeling in vivo. *J. Clin. Investig.* **2006**, *116*, 163–173. [[CrossRef](#)] [[PubMed](#)]
103. Jing, H.; Yen, J.H.; Ganea, D. A novel signaling pathway mediates the inhibition of CCL3/4 expression by prostaglandin E2. *J. Biol. Chem.* **2004**, *279*, 55176–55186. [[CrossRef](#)] [[PubMed](#)]
104. Zhu, X.; Wei, D.; Chen, O.; Zhang, Z.; Xue, J.; Huang, S.; Zhu, W.; Wang, Y. Upregulation of CCL3/MIP-1alpha regulated by MAPKs and NF-kappaB mediates microglial inflammatory response in LPS-induced brain injury. *Acta Neurobiol. Exp.* **2016**, *76*, 304–317. [[CrossRef](#)]
105. Wang, J.; Tian, Y.; Phillips, K.L.; Chiverton, N.; Haddock, G.; Bunning, R.A.; Cross, A.K.; Shapiro, I.M.; Le Maitre, C.L.; Risbud, M.V. Tumor necrosis factor alpha- and interleukin-1beta-dependent induction of CCL3 expression by nucleus pulposus cells promotes macrophage migration through CCR1. *Arthritis Rheum.* **2013**, *65*, 832–842. [[CrossRef](#)]
106. Kitanaka, N.; Nakano, R.; Sugiura, K.; Kitanaka, T.; Namba, S.; Konno, T.; Nakayama, T.; Sugiya, H. Interleukin-1beta promotes interleukin-6 expression via ERK1/2 signaling pathway in canine dermal fibroblasts. *PLoS ONE* **2019**, *14*, e0220262. [[CrossRef](#)]
107. Hu, S.L.; Huang, C.C.; Tzeng, T.T.; Liu, S.C.; Tsai, C.H.; Fong, Y.C.; Tang, C.H. S1P promotes IL-6 expression in osteoblasts through the PI3K, MEK/ERK and NF-kappaB signaling pathways. *Int. J. Med. Sci.* **2020**, *17*, 1207–1214. [[CrossRef](#)]
108. Klemm, C.; Bruchhagen, C.; van Kruchten, A.; Niemann, S.; Löffler, B.; Peters, G.; Ludwig, S.; Ehrhardt, C. Mitogen-activated protein kinases (MAPKs) regulate IL-6 over-production during concomitant influenza virus and Staphylococcus aureus infection. *Sci. Rep.* **2017**, *7*, 42473. [[CrossRef](#)]
109. Plastira, I.; Bernhart, E.; Joshi, L.; Koyani, C.N.; Strohmaier, H.; Reicher, H.; Malle, E.; Sattler, W. MAPK signaling determines lysophosphatidic acid (LPA)-induced inflammation in microglia. *J. Neuroinflamm.* **2020**, *17*, 127. [[CrossRef](#)]
110. Kojima, R.; Kawachi, M.; Ito, M. Butein suppresses ICAM-1 expression through the inhibition of I kappa B alpha and c-Jun phosphorylation in TNF-alpha- and PMA-treated HUVECs. *Int. Immunopharmacol.* **2015**, *24*, 267–275. [[CrossRef](#)]
111. Hadad, N.; Tuval, L.; Elgazar-Carmom, V.; Levy, R.; Levy, R. Endothelial ICAM-1 protein induction is regulated by cytosolic phospholipase A2alpha via both NF-kappaB and CREB transcription factors. *J. Immunol.* **2011**, *186*, 1816–1827. [[CrossRef](#)] [[PubMed](#)]
112. Jaramillo, M.; Olivier, M. Hydrogen peroxide induces murine macrophage chemokine gene transcription via extracellular signal-regulated kinase- and cyclic adenosine 5'-monophosphate (cAMP)-dependent pathways: Involvement of NF-kappa B, activator protein 1, and cAMP response element binding protein. *J. Immunol.* **2002**, *169*, 7026–7038. [[CrossRef](#)] [[PubMed](#)]
113. Salti, T.; Khazim, K.; Haddad, R.; Campisi-Pinto, S.; Bar-Sela, G.; Cohen, I. Glucose Induces IL-1alpha-Dependent Inflammation and Extracellular Matrix Proteins Expression and Deposition in Renal Tubular Epithelial Cells in Diabetic Kidney Disease. *Front. Immunol.* **2020**, *11*, 1270. [[CrossRef](#)] [[PubMed](#)]
114. Lei, Y.; Devarapu, S.K.; Motrapu, M.; Cohen, C.D.; Lindenmeyer, M.T.; Moll, S.; Kumar, S.V.; Anders, H.J. Interleukin-1beta Inhibition for Chronic Kidney Disease in Obese Mice With Type 2 Diabetes. *Front. Immunol.* **2019**, *10*, 1223. [[CrossRef](#)]

115. Bandach, I.; Segev, Y.; Landau, D. Experimental modulation of Interleukin 1 shows its key role in chronic kidney disease progression and anemia. *Sci. Rep.* **2021**, *11*, 6288. [[CrossRef](#)]
116. Stangou, M.; Papagianni, A.; Bantis, C.; Moisiadis, D.; Kasimatis, S.; Spartalis, M.; Pantzaki, A.; Efstratiadis, G.; Memmos, D. Up-regulation of urinary markers predict outcome in IgA nephropathy but their predictive value is influenced by treatment with steroids and azathioprine. *Clin. Nephrol.* **2013**, *80*, 203–210. [[CrossRef](#)]
117. Anders, H.J. Of Inflammasomes and Alarmins: IL-1beta and IL-1alpha in Kidney Disease. *J. Am. Soc. Nephrol.* **2016**, *27*, 2564–2575. [[CrossRef](#)]
118. Privratsky, J.R.; Zhang, J.; Lu, X.; Rudemiller, N.; Wei, Q.; Yu, Y.R.; Gunn, M.D.; Crowley, S.D. Interleukin 1 receptor (IL-1R1) activation exacerbates toxin-induced acute kidney injury. *Am. J. Physiol.* **2018**, *315*, F682–F691. [[CrossRef](#)]
119. Jones, L.K.; O'Sullivan, K.M.; Semple, T.; Kuligowski, M.P.; Fukami, K.; Ma, F.Y.; Nikolic-Paterson, D.J.; Holdsworth, S.R.; Kitching, A.R. IL-1RI deficiency ameliorates early experimental renal interstitial fibrosis. *Nephrol. Dial. Transplant.* **2009**, *24*, 3024–3032. [[CrossRef](#)]
120. Lavozy, C.; Matus, Y.S.; Orejudo, M.; Carpio, J.D.; Droguett, A.; Egado, J.; Mezzano, S.; Ruiz-Ortega, M. Interleukin-17A blockade reduces albuminuria and kidney injury in an accelerated model of diabetic nephropathy. *Kidney Int.* **2019**, *95*, 1418–1432. [[CrossRef](#)]
121. Orejudo, M.; Rodrigues-Diez, R.R.; Rodrigues-Diez, R.; Garcia-Redondo, A.; Santos-Sanchez, L.; Rande-Garbayo, J.; Cannata-Ortiz, P.; Ramos, A.M.; Ortiz, A.; Selgas, R.; et al. Interleukin 17A Participates in Renal Inflammation Associated to Experimental and Human Hypertension. *Front. Pharmacol.* **2019**, *10*, 1015. [[CrossRef](#)] [[PubMed](#)]
122. Mehrotra, P.; Collett, J.A.; McKinney, S.D.; Stevens, J.; Ivancic, C.M.; Basile, D.P. IL-17 mediates neutrophil infiltration and renal fibrosis following recovery from ischemia reperfusion: Compensatory role of natural killer cells in athymic rats. *Am. J. Physiol.* **2017**, *312*, F385–F397. [[CrossRef](#)] [[PubMed](#)]
123. Sun, B.; Wang, H.; Zhang, L.; Yang, X.; Zhang, M.; Zhu, X.; Ji, X.; Wang, H. Role of interleukin 17 in TGF-beta signaling-mediated renal interstitial fibrosis. *Cytokine* **2018**, *106*, 80–88. [[CrossRef](#)] [[PubMed](#)]
124. Ramani, K.; Tan, R.J.; Zhou, D.; Coleman, B.M.; Jawale, C.V.; Liu, Y.; Biswas, P.S. IL-17 Receptor Signaling Negatively Regulates the Development of Tubulointerstitial Fibrosis in the Kidney. *Mediat. Inflamm.* **2018**, *2018*, 5103672. [[CrossRef](#)] [[PubMed](#)]
125. Sanz, A.B.; Justo, P.; Sanchez-Nino, M.D.; Blanco-Colio, L.M.; Winkles, J.A.; Kretzler, M.; Jakubowski, A.; Blanco, J.; Egado, J.; Ruiz-Ortega, M.; et al. The cytokine TWEAK modulates renal tubulointerstitial inflammation. *J. Am. Soc. Nephrol.* **2008**, *19*, 695–703. [[CrossRef](#)]
126. Ortiz, A.; Sanz, A.B.; Munoz Garcia, B.; Moreno, J.A.; Sanchez Nino, M.D.; Martin-Ventura, J.L.; Egado, J.; Blanco-Colio, L.M. Considering TWEAK as a target for therapy in renal and vascular injury. *Cytokine Growth Factor Rev.* **2009**, *20*, 251–258. [[CrossRef](#)]
127. Sanz, A.B.; Izquierdo, M.C.; Sanchez-Nino, M.D.; Ucerro, A.C.; Egado, J.; Ruiz-Ortega, M.; Ramos, A.M.; Putterman, C.; Ortiz, A. TWEAK and the progression of renal disease: Clinical translation. *Nephrol. Dial. Transplant.* **2014**, *29* (Suppl. S1), i54–i62. [[CrossRef](#)]
128. Ucerro, A.C.; Benito-Martin, A.; Fuentes-Calvo, I.; Santamaria, B.; Blanco, J.; Lopez-Novoa, J.M.; Ruiz-Ortega, M.; Egado, J.; Burkly, L.C.; Martinez-Salgado, C.; et al. TNF-related weak inducer of apoptosis (TWEAK) promotes kidney fibrosis and Ras-dependent proliferation of cultured renal fibroblast. *Biochim. Biophys. Acta* **2013**, *1832*, 1744–1755. [[CrossRef](#)]
129. Bernardi, S.; Voltan, R.; Rimondi, E.; Melloni, E.; Milani, D.; Cervellati, C.; Gemmati, D.; Celeghini, C.; Secchiero, P.; Zauli, G.; et al. TRAIL, OPG, and TWEAK in kidney disease: Biomarkers or therapeutic targets? *Clin. Sci.* **2019**, *133*, 1145–1166. [[CrossRef](#)]
130. Sun, F.; Teng, J.; Yu, P.; Li, W.; Chang, J.; Xu, H. Involvement of TWEAK and the NF-kappaB signaling pathway in lupus nephritis. *Exp. Ther. Med.* **2018**, *15*, 2611–2619. [[CrossRef](#)]
131. Poveda, J.; Tabara, L.C.; Fernandez-Fernandez, B.; Martin-Cleary, C.; Sanz, A.B.; Selgas, R.; Ortiz, A.; Sanchez-Nino, M.D. TWEAK/Fn14 and Non-Canonical NF-kappaB Signaling in Kidney Disease. *Front. Immunol.* **2013**, *4*, 447. [[CrossRef](#)] [[PubMed](#)]
132. Sanz, A.B.; Sanchez-Nino, M.D.; Ortiz, A. TWEAK, a multifunctional cytokine in kidney injury. *Kidney Int.* **2011**, *80*, 708–718. [[CrossRef](#)] [[PubMed](#)]
133. Mehaffey, E.; Majid, D.S.A. Tumor necrosis factor-alpha, kidney function, and hypertension. *Am. J. Physiol.* **2017**, *313*, F1005–F1008. [[CrossRef](#)]
134. Vielhauer, V.; Mayadas, T.N. Functions of TNF and its receptors in renal disease: Distinct roles in inflammatory tissue injury and immune regulation. *Semin. Nephrol.* **2007**, *27*, 286–308. [[CrossRef](#)] [[PubMed](#)]
135. Hernandez, T.; Mayadas, T.N. Immunoregulatory role of TNFalpha in inflammatory kidney diseases. *Kidney Int.* **2009**, *76*, 262–276. [[CrossRef](#)]
136. Nikitopoulou, I.; Oikonomou, N.; Karouzakis, E.; Sevastou, I.; Nikolaidou-Katsaridou, N.; Zhao, Z.; Mersinias, V.; Armaka, M.; Xu, Y.; Masu, M.; et al. Autotaxin expression from synovial fibroblasts is essential for the pathogenesis of modeled arthritis. *J. Exp. Med.* **2012**, *209*, 925–933. [[CrossRef](#)]
137. Lu, J.; Chatterjee, M.; Schmid, H.; Beck, S.; Gawaz, M. CXCL14 as an emerging immune and inflammatory modulator. *J. Inflamm.* **2016**, *13*, 1. [[CrossRef](#)]
138. Lv, J.; Wu, Z.L.; Gan, Z.; Gui, P.; Yao, S.L. CXCL14 Overexpression Attenuates Sepsis-Associated Acute Kidney Injury by Inhibiting Proinflammatory Cytokine Production. *Mediat. Inflamm.* **2020**, *2020*, 2431705. [[CrossRef](#)]

139. Mirzoyan, K.; Baiotto, A.; Dupuy, A.; Marsal, D.; Denis, C.; Vinel, C.; Sicard, P.; Bertrand-Michel, J.; Bascands, J.L.; Schanstra, J.P.; et al. Increased urinary lysophosphatidic acid in mouse with subtotal nephrectomy: Potential involvement in chronic kidney disease. *J. Physiol. Biochem.* **2016**, *72*, 803–812. [[CrossRef](#)]
140. Swaney, J.S.; Chapman, C.; Correa, L.D.; Stebbins, K.J.; Broadhead, A.R.; Bain, G.; Santini, A.M.; Darlington, J.; King, C.D.; Baccei, C.S.; et al. Pharmacokinetic and pharmacodynamic characterization of an oral lysophosphatidic acid type 1 receptor-selective antagonist. *J. Pharmacol. Exp. Ther.* **2011**, *336*, 693–700. [[CrossRef](#)]
141. Grove, K.J.; Voziyan, P.A.; Spraggins, J.M.; Wang, S.; Pauksakon, P.; Harris, R.C.; Hudson, B.G.; Caprioli, R.M. Diabetic nephropathy induces alterations in the glomerular and tubule lipid profiles. *J. Lipid Res.* **2014**, *55*, 1375–1385. [[CrossRef](#)] [[PubMed](#)]
142. Li, H.Y.; Oh, Y.S.; Choi, J.W.; Jung, J.Y.; Jun, H.S. Blocking lysophosphatidic acid receptor 1 signaling inhibits diabetic nephropathy in db/db mice. *Kidney Int.* **2017**, *91*, 1362–1373. [[CrossRef](#)] [[PubMed](#)]
143. Mizuno, S.; Matsumoto, K.; Nakamura, T. HGF as a renoprotective and anti-fibrotic regulator in chronic renal disease. *Front. Biosci.* **2008**, *13*, 7072–7086. [[CrossRef](#)] [[PubMed](#)]
144. Nishino, T.; Pusey, C.D.; Domin, J. Elevated Akt phosphorylation as an indicator of renal tubular epithelial cell stress. *J. Biol. Chem.* **2002**, *277*, 33943–33949. [[CrossRef](#)]
145. Veerasamy, M.; Nguyen, T.Q.; Motazed, R.; Pearson, A.L.; Goldschmeding, R.; Dockrell, M.E. Differential regulation of E-cadherin and alpha-smooth muscle actin by BMP 7 in human renal proximal tubule epithelial cells and its implication in renal fibrosis. *Am. J. Physiol.* **2009**, *297*, F1238–F1248. [[CrossRef](#)]
146. Zeisberg, M.; Hanai, J.; Sugimoto, H.; Mammoto, T.; Charytan, D.; Strutz, F.; Kalluri, R. BMP-7 counteracts TGF-beta1-induced epithelial-to-mesenchymal transition and reverses chronic renal injury. *Nat. Med.* **2003**, *9*, 964–968. [[CrossRef](#)]
147. Moll, S.; Ebeling, M.; Weibel, F.; Farina, A.; Araujo Del Rosario, A.; Hoflack, J.C.; Pomposiello, S.; Prunotto, M. Epithelial cells as active player in fibrosis: Findings from an in vitro model. *PLoS ONE* **2013**, *8*, e56575. [[CrossRef](#)]
148. Broadbelt, N.V.; Chen, J.; Silver, R.B.; Poppas, D.P.; Felsen, D. Pressure activates epidermal growth factor receptor leading to the induction of iNOS via NFkappaB and STAT3 in human proximal tubule cells. *Am. J. Physiol.* **2009**, *297*, F114–F124. [[CrossRef](#)]
149. Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS A J. Integr. Biol.* **2012**, *16*, 284–287. [[CrossRef](#)]
150. Davis, S.; Meltzer, P.S. GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **2007**, *23*, 1846–1847. [[CrossRef](#)]
151. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [[CrossRef](#)] [[PubMed](#)]
152. Carvalho, B.S.; Irizarry, R.A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **2010**, *26*, 2363–2367. [[CrossRef](#)] [[PubMed](#)]
153. Kauffmann, A.; Gentleman, R.; Huber, W. arrayQualityMetrics—A bioconductor package for quality assessment of microarray data. *Bioinformatics* **2009**, *25*, 415–416. [[CrossRef](#)] [[PubMed](#)]
154. Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck, W.M., III; Hao, Y.; Stoeckius, M.; Smibert, P.; Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177*, 1888–1902.e1821. [[CrossRef](#)] [[PubMed](#)]
155. Cabello-Aguilar, S.; Alame, M.; Kon-Sun-Tack, F.; Fau, C.; Lacroix, M.; Colinge, J. SingleCellSignalR: Inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* **2020**, *48*, e55. [[CrossRef](#)] [[PubMed](#)]
156. Shao, X.; Liao, J.; Li, C.; Lu, X.; Cheng, J.; Fan, X. CellTalkDB: A manually curated database of ligand–receptor interactions in humans and mice. *Brief. Bioinform.* **2021**, *22*, bbaa269. [[CrossRef](#)] [[PubMed](#)]
157. Gu, Z.; Gu, L.; Eils, R.; Schlesner, M.; Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **2014**, *30*, 2811–2812. [[CrossRef](#)]