



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ – ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΛΟΓΙΑ



ΕΛΛΗΝΙΚΟ ΙΝΣΤΙΤΟΥΤΟ ΠΑΣΤΕΡ
INSTITUT PASTEUR HELLENIQUE

Διπλωματική Εργασία



«Ανάπτυξη βιοπληροφορικών ροών και εργαλείων για την μελέτη της ανοσοβιολογικής απόκρισης στον SARS-CoV-2, με βάση τις τεχνολογίες αλληλούχισης επόμενης γενεάς (NGS–RNAseq)».

ΡΕΠΟΥΣΗ ΝΙΚΟΛΕΝΑ

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ

1. Ιωάννης Τρουγκάκος, Καθηγητής, Τμήμα Βιολογίας, ΕΚΠΑ (Επιβλέπων)
2. Βασιλική Οικονομίδου, Αναπληρώτρια Καθηγήτρια, Τμήμα Βιολογίας, ΕΚΠΑ
3. Τιμοκράτης Καραμήτρος, Εντεταλμένος Ερευνητής, Μονάδα Βιοπληροφορικής και Εφαρμοσμένης Γενομικής, Ελληνικό Ινστιτούτο Παστέρ

ΑΘΗΝΑ

ΝΟΕΜΒΡΙΟΣ 2021

**Αφιερωμένο στο Ίδρυμα Παιδείας
και Ευρωπαϊκού Πολιτισμού**

ΕΥΧΑΡΙΣΤΙΕΣ

Φτάνοντας στο τέλος των μεταπτυχιακών μου σπουδών, και ολοκληρώνοντας τη μεταπτυχιακή διπλωματική μου εργασία, θα ήθελα να ευχαριστήσω όλους τους ανθρώπους, που με βοήθησαν.

Ειδικότερα, θα ήθελα να ευχαριστήσω:

- Τα μέλη της τριμελούς μου επιτροπής, Καθηγητή Ιωάννη Τρουγκάκο, Δρ Τιμοκράτη Καραμήτρο και Αναπληρώτρια Καθηγήτρια Βασιλική Οικονομίδου. Θα ήθελα να ευχαριστήσω ιδιαίτερα τον Δρ Τιμοκράτη Καραμήτρο για το χρόνο του.
- Το Ίδρυμα Παιδείας και Ευρωπαϊκού Πολιτισμού και το Κληροδότημα Σπυρίδωνος Αντύπα Υπέρ Κεφαλληνίας, για την οικονομική στήριξη των μεταπτυχιακών μου σπουδών. Θα ήθελα να ευχαριστήσω θερμά τον Καθηγητή Ιωάννη Τρουγκάκο, την Καθηγήτρια Ισιδώρα Παπασιδέρη και τη γραμματέα Ευαγγελία Παπατρέχα για τη βοήθεια στην εξασφάλιση της υποτροφίας.
- Τους συμφοιτητές μου και τους συνοδοιπόρους μου στο εργαστήριο, και κυρίως τη Γεσθημανή και την Ελένη.
- Τη Δρ Βασιλική Πόγκα, μεταδιδακτορική ερευνήτρια στο Ελληνικό Ινστιτούτο Παστέρ, για τη βοήθεια στο πειραματικό μέρος της παρούσας εργασίας.
- Την οικογένεια μου και τους φίλους μου, και ιδιαίτερα τον αδερφό μου Νικηφόρο για τις πολύτιμες συμβουλές του και τη μητέρα μου Σίσσυ, για τη βοήθεια.

Νικολένα Ρεπούση, Νοέμβριος 2021

ΠΕΡΙΛΗΨΗ

Ο SARS-CoV-2 είναι ένας νέος βήτα κορωνοϊός, ο οποίος εμφανίστηκε στο τέλος του 2019, στην επαρχία Hubei της Κίνας. Προκαλεί τη νόσο coronavirus disease 2019 (COVID-19), η οποία ανακηρύχθηκε ως πανδημία από τον Παγκόσμιο Οργανισμό Υγείας (ΠΟΥ) στις 11 Μαρτίου του 2020. Η κλινική εικόνα της νόσου COVID-19 εξαρτάται σε μεγάλο βαθμό από την ανοσολογική απόκριση. Στην ανοσολογική απόκριση στη COVID-19 συμμετέχει τόσο η μη ειδική ανοσία, όσο και η ειδική ανοσία - χυμική και κυτταρική-. Έπειτα από την είσοδο του SARS-CoV-2 στα ανθρώπινα κύτταρα, ιογενή RNAs απελευθερώνονται και δρουν σαν ειδικά μοριακά πρότυπα των παθογόνων (Pathogen-associated molecular patterns, PAMPs), τα οποία αναγνωρίζονται από υποδοχείς αναγνώρισης προτύπων (Pattern recognition receptors, PRRs), όπως ανάλογοι των Toll υποδοχείς (Toll-like receptors, TLRs). Σαν αποτέλεσμα, μεταξύ άλλων, εκφράζονται προφλεγμονώδεις κυτταροκίνες, χημειοκίνες (πχ, IL-1, IL-6) και διαλυτοί παράγοντες που συσχετίζονται με γονίδια που κωδικοποιούν ιντερφερόνες (Interferon-Stimulated Genes, ISGs). Ακόμη, μέρη του ιού, όπως η πρωτεΐνη-ακίδα, μπορούν να αναγνωριστούν από τα Β λεμφοκύτταρα και να παρουσιαστούν από MHC συμπλέγματα στα Τ λεμφοκύτταρα, με επακόλουθο την παραγωγή αντισωμάτων και την κυτταρολυτική δραστηριότητα κατά την οξεία φάση της λοίμωξης. Η αλληλούχιση του RNA (RNA-SEQ) και η εξειδικευμένη βιοπληροφορική ανάλυση του μεταγραφώματος ασθενών με COVID19 αποτελεί βασικό εργαλείο μελέτης της διαφορικής έκφρασης των γονιδίων μεταξύ ομάδων ασθενών. Η γενική ροή εργασίας κατά την ανάλυση της γονιδιακής έκφρασης αποτελείται από το πειραματικό μέρος και το υπολογιστικό μέρος. Το πειραματικό μέρος συνίσταται από τα ακόλουθα βήματα: απομόνωση RNA, προετοιμασία βιβλιοθήκης, έλεγχος ποιότητας και αλληλούχιση επόμενης γενιάς, ενώ η βιοπληροφορική ανάλυση συνίσταται από τα εξής βήματα: επεξεργασία των αναγνώσεων (συμπεριλαμβανομένης της στοίχισης, alignment), εκτίμηση των επιμέρους επιπέδων γονιδιακής έκφρασης, κανονικοποίηση, ταυτοποίηση γονιδίων με ανάλυση διαφορικής έκφρασης. Το QuantSeq αποτελεί μία μέθοδο προετοιμασίας δείγματος RNA με σκοπό τον

ακριβή προσδιορισμό της γονιδιακής έκφρασης. Το QuantSeq παρέχει ένα ευέλικτο πρωτόκολλο για τη δημιουργία βιβλιοθηκών αλληλούχισης επόμενης γενιάς με υψηλή εξειδίκευση στην αλυσίδα κοντά στο 3' άκρο πολυαδενυλιωμένων RNA. Παράγεται μόνο ένα θραύσμα ανά μεταγραφή, συνδέοντας άμεσα τον αριθμό των αναγνώσεων που αντιστοιχούν σε ένα γονίδιο με την έκφρασή του.

Στη συγκεκριμένη εργασία, εφαρμόστηκε το πρωτόκολλο QuantSeq για τη μελέτη της ανοσοβιολογικής απόκρισης 8 ασθενών με COVID-19, οι οποίοι διέφεραν ως προς το χρόνο αρνητικοποίησης του RT-PCR SARS-CoV-2 test. Ειδικότερα, έγινε ανάλυση διαφορικής έκφρασης μεταξύ 4 ασθενών με νόσο COVID-19, οι οποίοι είχαν θετικό RT-PCR SARS-CoV-2 test τουλάχιστον ένα μήνα μετά την αρχική θετικοποίηση, και 4 ασθενών που είχαν αρνητικό δεύτερο test μία εβδομάδα μετά την αρχική θετικοποίηση. Για το σκοπό αυτό, μετά την εκτέλεση του πειράματος αλληλούχισης QuantSeq, πραγματοποιήθηκε η βιοπληροφορική ανάλυση με τη χρήση διαφόρων εργαλείων για το κάθε βήμα, το BBduk για το trimming, το STAR και το Salmon για τη στοίχιση, το DESeq2 και το edgeR για την ανάλυση διαφορικής έκφρασης, το DAVID και το g: Profiler για τη λειτουργική ανάλυση και τα CIBERSORTx και xCell για την ψηφιακή κυτταρομετρία ροής. Η βιοπληροφορική ροή που επιλέχθηκε ήταν η κάτωθι: STAR \Rightarrow DESeq2 \Rightarrow DAVID. Από την ανάλυση διαφορικής έκφρασης βρέθηκαν 4464 διαφορεικά εκφρασμένα γονίδια σε επίπεδο σημαντικότητας 5% και $|\log_2\text{foldchange}| > 1$. Επίσης, από τη λειτουργική ανάλυση, προέκυψε ότι οι γονιδιακές οντολογίες που εμπλουτίστηκαν στατιστικά σημαντικά ήταν οι εξής: κυτταρική προσκόλληση (cell adhesion), χημική συναπτική μετάδοση (chemical synaptic transmission), σηματοδότηση μεσολαβούμενη από ιντερφερόνη γάμμα (interferon-gamma-mediated) και τα κάτωθι μονοπάτια: μονοπάτι σηματοδότησης ασβεστίου (calcium signaling pathway), εθισμός στη μορφίνη (morphine addiction), καθοδήγηση του άξονα των νευρικών κυττάρων (axon guidance), μόρια κυτταρικής προσκόλλησης (cell adhesion molecules, CAMs), διαφοροποίηση οστεοκλαστών (osteoclast differentiation).

ABSTRACT

SARS-CoV-2 is a new beta coronavirus, which appeared in late 2019, in Hubei Province, China. It causes coronavirus disease 2019 (COVID-19), which was declared a pandemic by the World Health Organization (WHO) on March 11, 2020. The clinical picture of COVID-19 is highly dependent on the immune response. Both non-specific immunity and specific immunity-humoral and cellular- are involved in the immune response to COVID-19. Following the entry of SARS-Cov-2 into human cells, viral RNAs are released and act as Pathogen-associated molecular patterns (PAMPs), which are recognized by Pattern recognition receptors (PRRs), such as Toll-like receptors (TLRs). As a result, among others, proinflammatory cytokines, chemokines (e.g., IL-1, IL-6) and soluble factors associated with Interferon-Stimulated Genes (ISGs) are expressed. Furthermore, parts of the virus, such as the spike protein, can be identified by B lymphocytes or presented by MHC complexes in T lymphocytes, resulting in the production of antibodies and cytolytic activity during the acute phase of the infection. RNA sequencing (RNA-SEQ) and specialized bioinformatics analysis of the transcriptome of patients with COVID-19 is a key tool for studying the differential expression of genes between groups of patients. The general workflow in a gene expression analysis consists of two parts, the experimental part and the computational part. The experimental part consists of the following steps: RNA extraction, library preparation, quality control and next generation sequencing, while the bioinformatics analysis consists of the following steps: reads mapping, reads quantification, normalization, gene identification by differential expression analysis. QuantSeq is an RNA sample preparation method for the precise determination of gene expression. QuantSeq provides an easy-to-use protocol for creating next-generation sequencing libraries with high strand specificity near the 3' end of polyadenylated RNA. Only one fragment is generated per transcript, directly linking the number of reads corresponding to a gene to its expression.

In this study, the QuantSeq protocol was applied to study the immune response of 8 patients with COVID-19, who differed in the duration of

positivity of their RT-PCR SARS-CoV-2 test. Specifically, differential expression analysis was performed between 4 patients with COVID-19 disease who tested positive for RT-PCR SARS-CoV-2 one month after initial positive, and 4 patients who tested negative, one week after the first positive test. For this purpose, after performing the QuantSeq sequencing experiment, bioinformatics analysis was performed using various tools for each step, BBduk for trimming, STAR and Salmon for alignment, DESeq2 and edgeR for differential expression analysis, DAVID and g: Profiler for functional analysis and CIBERSORTx and xCell for digital flow cytometry. The bioinformatics pipeline selected was the following one: STAR \Rightarrow DESeq2 \Rightarrow DAVID. Differential expression analysis found 4464 differentially expressed genes at a significance level of 5%, with $|\log_2(\text{foldchange})| > 1$. Also, the functional analysis, showed that the statistically significant enriched gene ontologies were: cell adhesion, chemical synaptic transmission, interferon-gamma-mediated and the pathways: calcium signaling pathway, morphine addiction, axon guidance, cell adhesion molecules (CAMs), osteoclast differentiation.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΕΡΙΛΗΨΗ	4
ABSTRACT	6
ΚΕΦΑΛΑΙΟ 1 – ΕΙΣΑΓΩΓΗ	9
1.1 ΥΠΟΚΕΦΑΛΑΙΟ 1 - SARS-CoV-2 και ΑΝΟΣΟΛΟΓΙΚΗ ΑΠΟΚΡΙΣΗ	9
1.2 ΥΠΟΚΕΦΑΛΑΙΟ 2 - ΑΛΛΗΛΟΥΧΙΣΗ	22
1.3 ΥΠΟΚΕΦΑΛΑΙΟ 3 - ΕΡΕΥΝΗΤΙΚΟ ΕΡΩΤΗΜΑ – ΣΚΟΠΟΣ ΤΗΣ ΕΡΓΑΣΙΑΣ	36
ΚΕΦΑΛΑΙΟ 2 - ΜΕΘΟΔΟΣ	37
2.1 ΔΕΙΓΜΑΤΑ	37
2.2 ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ	38
2.3 ΥΠΟΛΟΓΙΣΤΙΚΟ ΜΕΡΟΣ	40
ΚΕΦΑΛΑΙΟ 3 – ΑΠΟΤΕΛΕΣΜΑΤΑ	51
3.1 ΕΛΕΓΧΟΣ ΠΟΙΟΤΗΤΑΣ RNA	51
3.2. ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ΡΟΗ QUANTSEQ	52
ΚΕΦΑΛΑΙΟ 4 - ΣΥΖΗΤΗΣΗ	71
4.1 ΑΝΑΛΥΣΗ ΔΙΑΦΟΡΙΚΗΣ ΕΚΦΡΑΣΗΣ	72
4.2 ΛΕΙΤΟΥΡΓΙΚΗ ΑΝΑΛΥΣΗ	76
4.3. ΣΥΓΚΡΙΣΗ ΒΙΟΠΛΗΡΟΦΟΡΙΚΩΝ ΡΟΩΝ	78
ΚΕΦΑΛΑΙΟ 5 - ΣΥΜΠΕΡΑΣΜΑ – ΠΕΡΙΟΡΙΣΜΟΙ – ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΤΟ ΜΕΛΛΟΝ	80
5.1 ΣΥΜΠΕΡΑΣΜΑ	80
5.2 ΠΕΡΙΟΡΙΣΜΟΙ – ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΤΟ ΜΕΛΛΟΝ	81
ΚΕΦΑΛΑΙΟ 6 - ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	82
6.1 ΑΓΓΛΟΦΩΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ	82
6.2 ΕΛΛΗΝΟΦΩΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ	96

ΚΕΦΑΛΑΙΟ 1 – ΕΙΣΑΓΩΓΗ

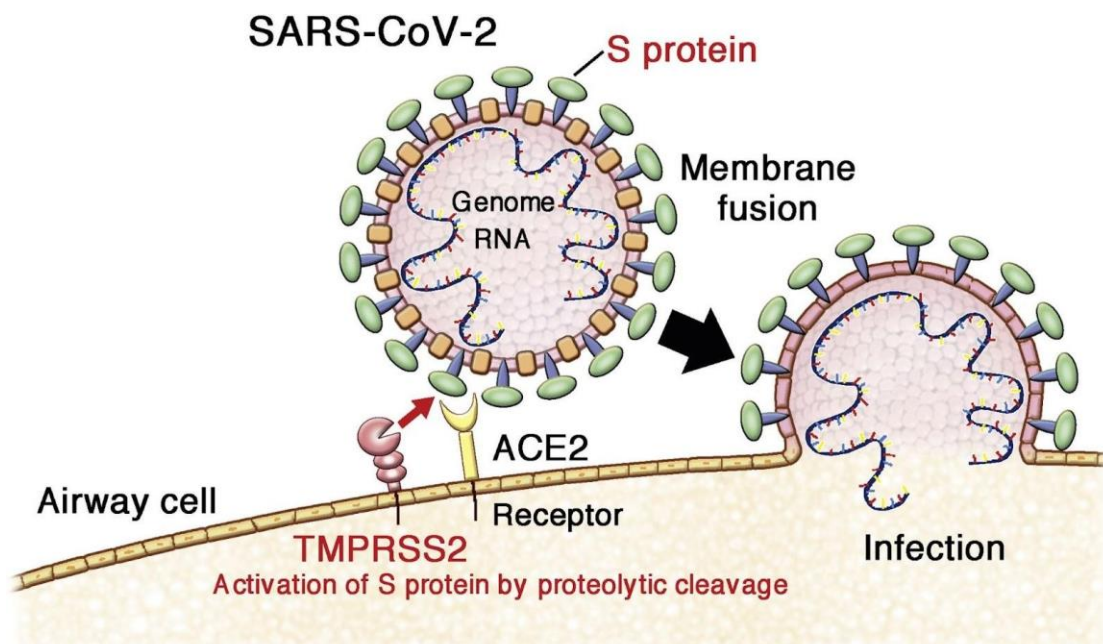
1.1 ΥΠΟΚΕΦΑΛΑΙΟ 1 - SARS-CoV-2 και ΑΝΟΣΟΛΟΓΙΚΗ ΑΠΟΚΡΙΣΗ

1.1.1 SARS-CoV-2

Ο Sars Coronavirus 2 (SARS-CoV-2) είναι ένα νέος βήτα κορωνοϊός, παρόμοιος με τον SARS-CoV-1, ο οποίος εμφανίστηκε στο τέλος του 2019, στην επαρχία Hubei της Κίνας. Προκαλεί τη νόσο coronavirus disease 2019 (COVID-19), η οποία ανακηρύχθηκε ως πανδημία από τον Παγκόσμιο Οργανισμό Υγείας (ΠΟΥ) στις 11 Μαρτίου του 2020 («Naming the coronavirus disease (COVID-19) and the virus that causes it». who.int). Ο SARS-CoV-2 είναι κορωνοϊός με γονιδίωμα μονόκλωνου RNA θετικής πολικότητας, είναι μεταδοτικός στον άνθρωπο και έχει γενετικές ομοιότητες με τους κορωνοϊούς SARS (SARS-CoV) (79,5%). Πιστεύεται ότι έχει προέλευση από τους ιούς που ευδοκιμούν σε ζώα και συγκεκριμένα στις νυχτερίδες με τις οποίες έχει υψηλή συσχέτιση (96%). Με βάση τις επιδημιολογικές μελέτες, κάθε μόλυνση από τον ιό έχει ως αποτέλεσμα κατά μέσο όρο 5,7 νέες μολύνσεις, όταν κανένα μέλος της κοινότητας δεν έχει ανοσία και δεν λαμβάνεται κανένα μέτρο πρόληψης από τη διασπορά του ιού. Ο ιός εξαπλώνεται μεταξύ των ανθρώπων μέσω της στενής επαφής και μέσω των αναπνευστικών σταγονιδίων που παράγονται κατά το βήχα ή το φτέρνισμα (van Doremalen et al., 2020, Machhi et al., 2020).

Δομικά, ο SARS-CoV-2 έχει τέσσερις κύριες δομικές πρωτεΐνες. Πιο συγκεκριμένα, ο SARS-CoV-2 έχει την πρωτεΐνη-ακίδα (spike glycoprotein, S), τη γλυκοπρωτεΐνη μικρού φακέλου (small envelope glycoprotein, E), τη μεμβρανική γλυκοπρωτεΐνη (membrane glycoprotein, M), και την πρωτεΐνη νουκλεοκαψιδίου (nucleocapsid protein, N), καθώς και αρκετές επικουρικές

πρωτεΐνες. Η πρωτεΐνη-ακίδα είναι μία διαμεμβρανική πρωτεΐνη περίπου 150 kDa που βρίσκεται στο εξωτερικό μέρος του ιού. Η πρωτεΐνη-ακίδα σχηματίζει ομοτριμερή, που είναι εκτεθειμένα στην ιική επιφάνεια και διευκολύνουν τη σύνδεση του ιού στα κύτταρα του ξενιστή, μέσω αλληλεπίδρασης της με το μετατρεπτικό ένζυμο της αγγειοτενσίνης 2 (angiotensin-converting enzyme 2, ACE2), που εκφράζεται, μεταξύ άλλων, και στα κύτταρα του κατώτερου αναπνευστικού συστήματος. Αυτή η γλυκοπρωτεΐνη αποικοδομείται από τα κύτταρα του ξενιστή σε δύο υπομονάδες, την S1 και την S2. Η υπομονάδα S1 είναι υπεύθυνη για τον καθορισμό του εύρους του ξενιστή και του κυτταρικού τροπισμού, ενώ η υπομονάδα S2 εμπλέκεται κυρίως στη σύντηξη του ιού με τα κύτταρα του ξενιστή. Η ένωση της πρωτεΐνης - ακίδας του SARS-CoV-2 με το διαμεμβρανικό υποδοχέα ACE-2 του ξενιστή, φαίνεται ότι επάγει τη διάσπαση της πρωτεΐνης μέσω αλληλεπίδρασης με κυτταρικές πρωτεάσες, όπως την TMPRSS2 και κατά συνέπεια επιτρέπει την είσοδο του ιού στο κύτταρο του ξενιστή (Εικόνα 1). Η πρωτεΐνη N συνδέεται με το RNA, και εμπλέκεται σε διαδικασίες που σχετίζονται με το γονιδίωμα του ιού, τον κύκλο ανιγραφής του ιού και την κυτταρική απόκριση των κυττάρων ξενιστών σε ιογενείς λοιμώξεις (Heininger, 2020).

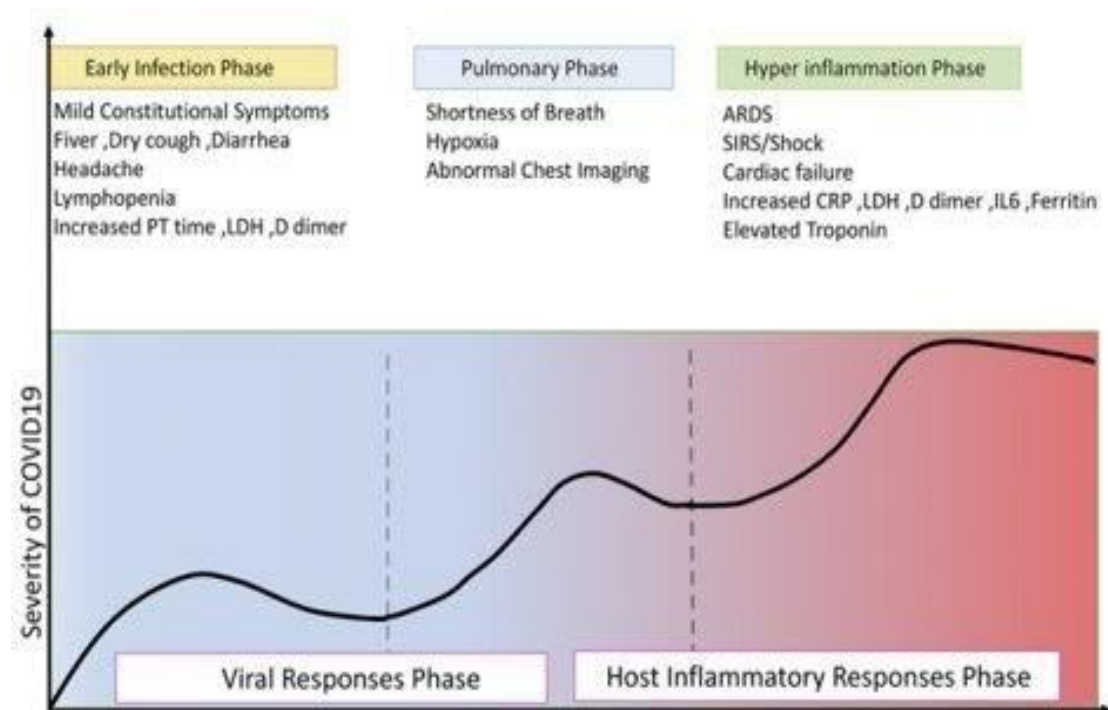


Εικόνα 1. Είσοδος του ιού SARS-CoV-2 σε κύτταρο του ξενιστή

1.1.2 COVID-19

Η νόσος COVID-19 είναι η νόσος που προκαλείται από τον κορωνοϊό SARS-CoV-2. Η νόσος COVID-19 ποικίλλει σημαντικά ως προς τη σοβαρότητά της, ανάλογα με τη φάση της λοίμωξης και της ανοσολογικής απόκρισης (Εικόνα 2). Μπορεί να υπάρχει πλήρης απουσία συμπτωμάτων (ασυμπτωματικοί ασθενείς) ή να εμφανιστούν πυρετός, βήχας, φαρυγγαλγία, αλλαγή ή απώλεια γεύσης ή/και οσμής, γενικότερη αδυναμία, διάρροια, κόπωση, μυϊκός πόνος. Σε σοβαρές περιπτώσεις, τα συμπτώματα μπορεί να περιλαμβάνουν, σοβαρή λοίμωξη των πνευμόνων, σύνδρομο οξείας αναπνευστικής δυσχέρειας (Acute Respiratory Distress Syndrome, ARDS), σήψη και σηπτική καταπληξία — γενικευμένη λοίμωξη και φλεγμονώδη αντίδραση, τα οποία χρήζουν εξειδικευμένης ιατρικής φροντίδας και υποστήριξης. Η κατάσταση ενός ασθενούς μπορεί να επιδεινωθεί ταχέως, συχνά κατά τη διάρκεια της δεύτερης εβδομάδας της νόσου. Ένα ποσοστό των ασθενών που θα προσβληθούν από COVID-19 θα χρειαστούν νοσηλεία, ορισμένοι ακόμη και παραμονή σε μονάδα εντατικής θεραπείας, ενίοτε για μεγάλα χρονικά διαστήματα. Οι ασθενείς που εμφανίζουν σοβαρά αναπνευστικά συμπτώματα μπορεί να χρειαστούν μηχανική υποστήριξη της αναπνοής, γεγονός που μπορεί να τους καταστήσει πιο ευάλωτους σε δευτερογενείς βακτηριακές λοιμώξεις. Επιπλέον, οι ασθενείς με COVID-19 μπορεί να παρουσιάσουν επιπλοκές, όπως εγκεφαλικό επεισόδιο ή καρδιακή προσβολή, λόγω της αυξημένης πηκτικότητας του αίματος. Ορισμένοι ασθενείς εμφανίζουν νευρολογικά συμπτώματα, όπως αλλαγές στην προσωπικότητα ή αλλαγές στην εγρήγορση (COVID-19: Clinical Features - UpToDate, n.d.). Η μακροχρόνιος COVID-19 (long COVID-19), γνωστή και ως σύνδρομο μετά COVID (post COVID syndrome), post-acute sequelae of COVID-19 (PASC), long-haul COVID και χρόνιο σύνδρομο COVID (chronic COVID syndrome, CCS) είναι μία νόσος για την οποία τα δεδομένα είναι λιγοστά. Η μακροχρόνιος COVID-19 χαρακτηρίζεται από μακροχρόνιες συνέπειες που επιμένουν μετά την τυπική περίοδο ανάρρωσης της COVID-19 (Nalbandian et al., 2021). Εκτιμάται ότι το 10% των επιζώντων από νόσο COVID-19 αναπτύσσουν long COVID-19 και για μεγάλο χρονικό διάστημα βιώνουν

συμπτώματα όπως βήχας, δύσπνοια, πόνος στο στήθος, πονοκέφαλος, δυσκολία συγκέντρωσης, αρθραλγία, μυϊκούς πόνους, κόπωση και διαταραχές της όσφρησης (Bertholee et al., 2017, Trisha Greenhalgh, Matthew Knight, Christine A'Court, Maria Buxton, 2020). Εκτός από τη σοβαρότητα της νόσου, άλλοι παράγοντες κινδύνου που συνδέονται με αυτό το σύνδρομο είναι η ηλικία και τα υποκείμενα νοσήματα. Είναι αξιοσημείωτο, ότι νεαρά άτομα, χωρίς συννοσηρότητες, παρόλο που νόσησαν ελαφριά, ανέπτυξαν long COVID-19 (S et al., 2021, Trisha Greenhalgh, Matthew Knight, Christine A'Court, Maria Buxton, 2020). Το πρώτο βήμα στην διαχείριση της νόσου COVID-19 είναι η γρήγορη και ακριβής ταυτοποίηση του ιού SARS-CoV-2 με τη χρήση αλυσιδωτής αντίδρασης πολυμεράσης σε πραγματικό χρόνο με αντίστροφη μεταγραφή (real-time reverse transcription-polymerase chain reaction, RT-PCR). Η RT-PCR ανιχνεύει το RNA του ιού SARS-CoV-2 σε ρινοφαρυγγικά δείγματα (Kevadiya et al., 2021).



Εικόνα 2. Κλινική εικόνα της νόσου COVID-19, ανάλογα με τη φάση της λοίμωξης και την ανοσολογική απόκριση

1.1.3 ΑΝΟΣΟΛΟΓΙΚΗ ΑΠΟΚΡΙΣΗ ΣΤΟΥΣ ΙΟΥΣ

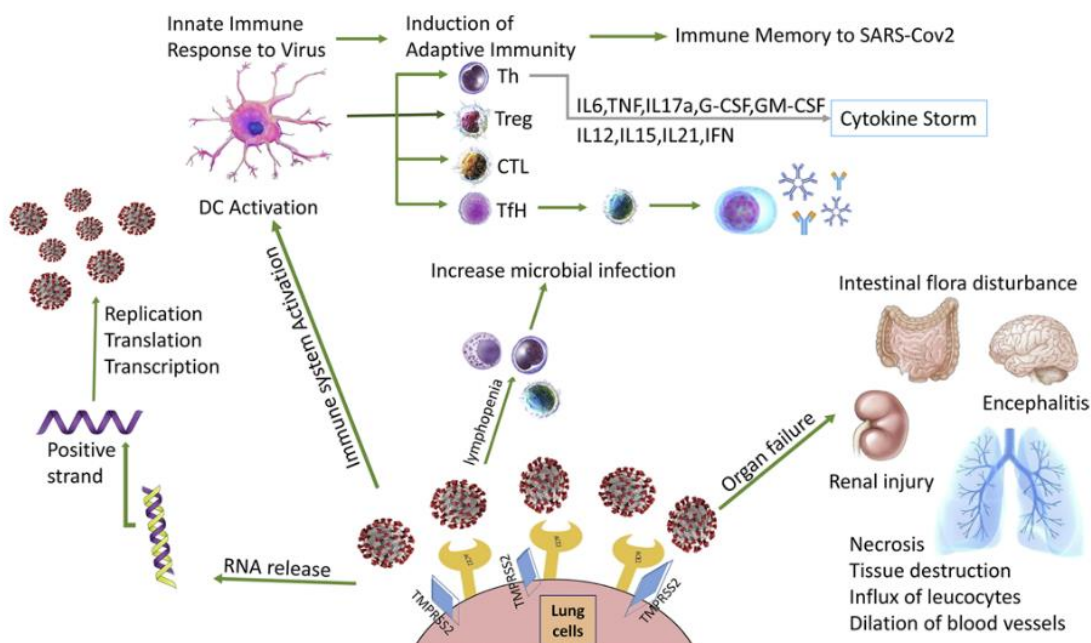
Το ανοσοποιητικό σύστημα είναι η φυσική ικανότητα του οργανισμού να αμύνεται έναντι παθογόνων (ιοί, βακτήρια) και να αντιμετωπίζει τις μολύνσεις. Υπάρχουν δύο τύποι ανοσίας, η ειδική και η μη ειδική ανοσία. Η μη ειδική ανοσία αποτελεί την πρώτη γραμμή άμυνας, και τη γενικευμένη άμεση απόκριση σε κάθε είδος παθογόνου. Αφού ένας ιός εισέλθει στον οργανισμό, τα ειδικά μοριακά πρότυπα των παθογόνων (pathogen-associated molecular patterns, PAMPs) του ιού τυπικά ανιχνεύονται από υποδοχείς αναγνώρισης προτύπων (pattern recognition receptors, PRRs) όπως οι ανάλογοι των Toll υποδοχείς (Toll-like receptors, TLRs), οι οποίοι σηματοδοτούν την ενεργοποίηση της μη ειδικής ανοσίας. Αυτό περιλαμβάνει τη στρατολόγηση, ενεργοποίηση και διαφοροποίηση κυττάρων του ανοσοποιητικού, συμπεριλαμβανομένων των ουδετερόφιλων, δενδριτικών κυττάρων, μακροφάγων, CD8 T κυττάρων και κυττάρων φυσικοί φονείς (Natural Killers, NK) και την απελευθέρωση ιντερφερονών (IFNs) και φλεγμονωδών κυτταροκινών συμπεριλαμβανομένης της οικογένειας IL-1, IL-6 και TNF, που ενεργοποιούν μια τοπική και συστηματική απάντηση στη μόλυνση. Επίσης, η μη ειδική ανοσία ενεργοποιεί την ειδική ανοσολογική απόκριση. Η ειδική ανοσία αποτελεί τη δεύτερη γραμμή άμυνας. Ξεκινά 6 έως 8 ημέρες μετά τη λοίμωξη και περιλαμβάνει δύο τύπους λευκοκυττάρων, τα T κύτταρα (κυτταρική απόκριση) και τα B κύτταρα (αντισώματα). Αναφορικά με την ειδική ανοσία, τα T κύτταρα αναγνωρίζουν τα μολυσμένα από έναν συγκεκριμένο ιό κύτταρα και γρήγορα αυξάνονται σε αριθμό, για να αντιμετωπίσουν τη λοίμωξη. Τα T βοηθητικά κύτταρα CD4+ στρατολογούν άλλα κύτταρα του ανοσοποιητικού και διεγείρουν τα B κύτταρα να παράγουν αντισώματα ειδικά για τον ιό. Τα κυτταροτοξικά T κύτταρα CD8+ εξουδετερώνουν τα κύτταρα, στα οποία πολλαπλασιάζεται ο ιός και τα βοηθούν στην ύφεση ή την επίλυση της λοίμωξης (Mueller & Rouse, 2008).

Η επίλυση της λοίμωξης εξαρτάται σε μεγάλο βαθμό από την κυτταροτοξική δράση των CD8+ T κυττάρων και των NK κυττάρων, που καθιστούν δυνατή την κάθαρση των μολυσμένων με ιό κυττάρων. Η απομάκρυνση των

κυττάρων που έχουν μολυνθεί από ιό μπορεί να διευκολύνει μια υπερφλεγμονώδη κατάσταση που ονομάζεται σύνδρομο ενεργοποίησης Μακροφάγων (ΜΦ) ή «καταιγίδα κυτταροκινών» και τελικά βλάβη στα όργανα που μολύνονται (Ragab et al., 2020).

1.1.4 ΑΝΟΣΟΛΟΓΙΚΗ ΑΠΟΚΡΙΣΗ ΣΤΟΝ SARS-CoV-2

Η ανοσολογική απόκριση στον ιό SARS-CoV-2 διακρίνεται στη μη ειδική και στην ειδική. Η μη ειδική χαρακτηρίζεται από την παραγωγή ιντερφερονών και τη συμμετοχή κυττάρων της μυελικής σειράς, ενώ η ειδική διακρίνεται στην κυτταρική και στη χυμική ανοσία (Εικόνα 3).



Εικόνα 3. Σύνοψη ανοσολογικής απόκρισης στον ιό SARS-CoV-2

- **ΜΗ ΕΙΔΙΚΗ ΑΝΟΣΟΛΟΓΙΚΗ ΑΠΟΚΡΙΣΗ ΣΤΟΝ SARS-CoV-2**

Η αλληλεπίδραση του SARS-CoV-2 με τα κύτταρα-στόχους απαιτεί την αναγνώριση της πρωτεΐνης - ακίδας από τον υποδοχέα ACE2, και τη διάσπαση της πρωτεΐνης - ακίδας από τη διαμεμβρανική πρωτεΐνη σερίνης πρωτεάση 2 (TMPRSS2) του κυττάρου ξενιστή. Ο SARS-CoV-2 έτσι μπορεί

να αποκτήσει πρόσβαση στην ενδοκυτάρια οδό, που επιτρέπει στον ιό να εντοπίζεται μέσα στα ενδοσώματα, όπου το ιικό μονόκλωνο RNA αναγνωρίζεται από τους υποδοχείς των ενδοσωμάτων TLR 7 και 8. Σύντομα μετά την αναγνώριση του ιικού RNA, η πρωτεΐνη προσαρμογέας διαφοροποίησης μυελικής σειράς παράγοντας 88 (adapter protein myeloid differentiation factor 88, MyD88), είναι υπεύθυνη για την ενεργοποίηση και τη μεταφορά στον πυρήνα των μεταγραφικών παραγόντων, πυρηνικός παράγοντας κάππα ενισχυτής ελαφριάς αλυσίδας ενεργοποιημένων B κυττάρων (nuclear factor kappa-light-chain-enhancer of activated B cells, NF-κB) και ρυθμιστικός παράγοντας ιντερφερόνης-3 (interferon regulatory transcription factor, IRF-3) και 7 (interferon regulatory factor 7, IRF-7) (Vafaeinezhad et al., 2021). Αυτοί οι μεταγραφικοί παράγοντες βοηθούν στη μεταγραφή γονιδίων που κωδικοποιούν προφλεγμονώδεις κυτταροκίνες και ιντερφερόνες (IFNs) τύπου I και III. Επιπλέον, η πρωτεΐνη TLR3 μπορεί να ενεργοποιηθεί από το δίκλωνο (ds) RNA, που παράγεται κατά την αντιγραφή του ιού, συμβάλλοντας έτσι μέσω της οδού TRIF σε μια προστατευτική απόκριση στη μόλυνση από SARS-CoV-2. Η φωσφορυλίωση των IRF-3 και IRF-7 απαιτείται για την έκφραση των IFN (Mandal et al., 2020, Fortune et al., 2002, Ricci et al., 2021).

Η ιντερλευκίνη 6 (IL-6) έχει αναφερθεί ως πιθανός παθογόνος παράγοντας στην έναρξη του συνδρόμου οξείας αναπνευστικής δυσχέρειας (ARDS) σε ασθενείς με COVID-19. Η IL-6 έχει ποικίλες ανοσολογικές και βιολογικές δράσεις, συμπεριλαμβανομένης της διαφοροποίησης των κυττάρων του ανοσοποιητικού. Η IL-6 είναι επίσης κρίσιμη για την ανάπτυξη των κυττάρων του πλάσματος και την παραγωγή αντισωμάτων. Δεδομένα από ασθενείς με SARS-CoV-2 δείχνουν ότι η μεταγραφή του γονιδίου της IL-6 ξεκινά και διατηρείται έπειτα από τη μόλυνση του αναπνευστικού επιθηλίου. Ο ιός SARS-CoV-2 ενεργοποιεί τα μακροφάγα, με αποτέλεσμα την άφθονη και παρατεταμένη παραγωγή IL -6 που μπορεί να οδηγήσει στην καταιγίδα κυτταροκινών, σε βλάβη των ενδοθηλιακών κυττάρων, και στα κλινικά και παθολογικά χαρακτηριστικά του ARDS (Jordan, 2021).

Το ARDS και η υπερπηκτική κατάσταση, που παρατηρούνται σε ορισμένους ασθενείς με COVID-19 μπορούν να αποδοθούν στη συστηματική ενεργοποίηση του συμπληρώματος, μέσω της εναλλακτικής οδού, της κλασικής οδού και της συνδεδεμένης με τη λεκτίνη οδό. Έχει επίσης προταθεί ότι η πρωτεΐνη N του SARS-CoV-2 είναι ένας ισχυρός ενεργοποιητής της οδού MBL/SP και μπορεί να είναι υπεύθυνη για την ταχεία ανάπτυξη του ARDS σε ασθενείς που έχουν μολυνθεί με SARS-CoV-2 (Jordan, 2021).

❖ ΠΑΡΑΓΩΓΗ ΙΝΤΕΡΦΕΡΟΝΩΝ

Η ενεργοποίηση της μη ειδικής ανοσολογικής απόκρισης έναντι του ιού SARS-CoV-2 στο αρχικό στάδιο της μόλυνσης απαιτεί ισχυρή παραγωγή ιντερφερονών (IFN). Η οικογένεια IFN χωρίζεται σε τρεις κύριες ομάδες: τον τύπο I (κυρίως IFN- α/β), τον τύπο II (IFN γ) και τον τύπο III (IFN- λ). Συγκεκριμένα, οι IFN τύπου I εκκρίνονται από όλα τα εμπύρνηνα κύτταρα μετά από μόλυνση με τον ιό. Ο τύπος II IFN ή IFN γ παράγεται κυρίως από κύτταρα NK και από τα T βοηθητικά κύτταρα τύπου 1 (Th1) και παίζει κεντρικό ρόλο στη ρύθμιση της ειδικής και μη ειδικής ανοσίας. Οι IFN- λ έχουν κοινές λειτουργίες και ενδοκυτταρικές οδούς με τις IFN τύπου I.

Μόλις απελευθερωθούν από τα μολυσμένα κύτταρα, οι IFN τύπου I συνδέονται με τους διμερείς υποδοχείς IFN (IFNAR), που αποτελούνται από τις υπομονάδες IFNAR1 και IFNAR2, και, μέσω της ενεργοποίησης της κινάσης της τυροσίνης 2 και της κινάσης Janus 1, μεσολαβούν στη φωσφορυλίωση των μεταγραφικών παραγόντων STAT1 και STAT2. Οι φωσφορυλιωμένες STAT πρωτεΐνες ετεροδιμερίζονται με μόρια IRF-9 για να σχηματίσουν το σύμπλεγμα γονιδίου 3 που διεγείρεται από IFN, το οποίο, μετά τη μεταφορά του στον πυρήνα, ξεκινά τη μεταγραφή γονιδίων που διεγείρονται από IFN (Interferon-Stimulated Genes, ISGs) μέσω της σύνδεσής τους με στοιχεία απόκρισης που διεγείρονται από IFN (Interferon-sensitive response element, ISRE) εντός των γονιδιακών υποκινητών. Η τύπου III IFN, αν και συνδέεται με διαφορετικό επιφανειακό υποδοχέα που εκφράζεται κυρίως σε επιθηλιακά κύτταρα, μοιράζεται με την IFN τύπου I το ίδιο

ενδοκυτταρικό μονοπάτι. Επομένως, οι IFN τύπου I και III προστατεύουν τα κύτταρα ξενιστές προκαλώντας την έκφραση μιας πληθώρας πρωτεϊνών που κωδικοποιούν ISGs και παρεμβαίνουν στην αντιγραφή του ιού και, επομένως, περιορίζουν τόσο την εξάπλωση του ιού όσο και το ιικό φορτίο.

Το εξαιρετικά παθογόνο δυναμικό του SARS-CoV-2 βασίζεται κυρίως στην ιδιότυπη ικανότητα του να εμποδίζει την οδό IFN και, από την άλλη πλευρά, να προκαλεί αυξημένη παραγωγή προφλεγμονωδών χημειοκινών και κυτταροκινών, συγκεκριμένα ιντερλευκίνης 6 (IL-6), υποδεικνύοντας έτσι ότι ο SARS-CoV-2 έχει αναπτύξει πολλαπλούς αποτελεσματικούς μηχανισμούς για να μετατοπίσει την ισορροπία υπέρ μιας προφλεγμονώδους κατάστασης που παρεμβαίνει στην παραγωγή IFN. Αυτό συνεπάγεται επίσης ότι, με διαφορετικό τρόπο από τον SARS-CoV, ο SARS-CoV-2 αναπαράγεται πιο ενεργά και αποτελεσματικά στους ιστούς των πνευμόνων του ανθρώπου, όπου έχει βρεθεί πολύ υψηλότερο ιικό φορτίο πιθανώς λόγω ελαττωματικής ιικής κάθαρσης. Έχει βρεθεί ότι χαμηλά επίπεδα IFN-I και ISGs δεν συσχετίζονται με υψηλότερο ιικό φορτίο στο αίμα και φλεγμονώδη απόκριση σε ασθενείς με σοβαρή νόσηση από COVID-19 σε σύγκριση με ήπιες περιπτώσεις. Η αυξημένη έκφραση ISGs σε μονοπύρηννα κύτταρα περιφερικού αίματος (Peripheral Blood Mononuclear Cell, PBMC), σε ασθενείς με ήπια ή ασυμπτωματική νόσηση από COVID-19 έναντι ασθενών με σοβαρή νόσηση, πιθανόν να εξαρτάται από την πρώιμη και ισχυρή παραγωγή IFN στους πνεύμονες που στη συνέχεια διαχέονται στην κυκλοφορία του αίματος. Επιπρόσθετα, έχει αναφερθεί ότι η μειωμένη έκφραση IFN τύπου I και τα υψηλά επίπεδα προφλεγμονωδών κυτταροκινών σε ορό ασθενών με σοβαρή COVID-19 αντιστρέφονται με τη βελτίωση της σοβαρότητας της νόσου και συσχετίζονται αντιστρόφως ανάλογα με το επίπεδο της IL-10. Όλα αυτά τα δεδομένα υποδεικνύουν ότι η αλληλεπίδραση μεταξύ της αντιγραφής του ιού και της έκφρασης των IFN παίζει κρίσιμο ρόλο στην πορεία της COVID-19 και θέτει τη βάση για το θεραπευτικό όφελος της χορήγησης IFN τύπου I σε επιλεγμένους ασθενείς νωρίς στην πορεία της λοίμωξης με SARS-CoV-2 (Pan et al., 2021, Choi & Shin, 2021).

❖ ΜΟΝΟΚΥΤΤΑΡΑ ΚΑΙ ΜΑΚΡΟΦΑΓΑ

Τα κύτταρα της μυελικής σειράς εμπλέκονται στην παθοφυσιολογία της μόλυνσης από τον κορωνοϊό SARS-CoV-2, είτε άμεσα, ως κύτταρα-στόχος του ιού, είτε έμμεσα, ως παραγωγοί προφλεγμονωδών κυτταροκινών. Το λεγόμενο σύνδρομο ενεργοποίησης μακροφάγων που χαρακτηρίζεται από υπερδραστικοποιημένα μακροφάγα έναντι του SARS-CoV-2 έχει συσχεπιστεί με ARDS. Αυτά τα στοιχεία συσχετίζονται με το εύρημα ότι ο SARS-CoV-2 ενεργοποιεί τα μακροφάγα μέσω του ACE2 και ενισχύει την έκκριση των IL-6, TNF-α και IL-10 (Knoll et al., 2021).

• ΕΙΔΙΚΗ ΑΝΟΣΟΛΟΓΙΚΗ ΑΠΟΚΡΙΣΗ ΣΤΟΝ SARS-CoV-2

Η ειδική ανοσολογική απόκριση διακρίνεται σε αυτή που μεσολαβείται από κύτταρα (κυτταρική ανοσία), και σε αυτή που μεσολαβείται από αντισώματα (χυμική ανοσία) (Or Caspi, Michael J. Smart, 2020).

❖ ΑΝΤΙΔΡΑΣΕΙΣ ΠΟΥ ΜΕΣΟΛΑΒΟΥΝΤΑΙ ΑΠΟ ΚΥΤΤΑΡΑ (ΚΥΤΤΑΡΙΚΗ ΑΝΟΣΙΑ)

Όταν ο ιός SARS-CoV-2 προσβάλλει τα αναπνευστικά επιθηλιακά κύτταρα του ανθρώπου, έπειτα από εισπνοή αεροσωματιδίων του ιού, τα δενδριτικά κύτταρα φαγοκυτταρώνουν τον ιό και παρουσιάζουν τμήματα του στα Τ κύτταρα. Τα Τ κύτταρα (effectors) λειτουργούν εξουδετερώνοντας τα μολυσμένα επιθηλιακά κύτταρα, και τα κυτταροτοξικά CD8+ Τ κύτταρα παράγουν και απελευθερώνουν προ-φλεγμονώδεις κυτταροκίνες που προκαλούν κυτταρική απόπτωση. Έχει αποδειχθεί ότι ασθενείς που μολύνθηκαν με SARS-CoV-2 έχουν υψηλά επίπεδα IL-1, IFN γ , IP-10 και MCP-1, που συνδέεται με αυξημένη Th1 ανοσολογική απόκριση. Όμως, η μόλυνση με SARS-CoV-2 συνδέεται επίσης με αυξημένη έκκριση κυτταροκινών Th2 (π.χ. IL-4 και IL-10) που καταστέλλουν τη φλεγμονή.

Ανάλυση κυτταρομετρίας ροής των PBMCs σε συμπτωματικούς ασθενείς με COVID-19 έχει δείξει σημαντικό αριθμό ενεργοποιημένων CD4+ T κυττάρων και μονοκυττάρων. Επιπλέον, σε σοβαρές περιπτώσεις COVID-19, έχει αναφερθεί μειωμένος αριθμός ρυθμιστικών T κυττάρων (κύτταρα Treg) και εντυπωσιακά υψηλός αριθμός κυττάρων CCR6+ Th17. Δεδομένου ότι τα κύτταρα Treg έχει αποδειχθεί ότι βοηθούν στην επίλυση της φλεγμονής ARDS, ο μειωμένος αριθμός κυττάρων Treg μπορεί να διευκολύνει την ανάπτυξη της ανοσοπαθολογίας της πνευμονικής νόσου COVID-19 (Ricci et al., 2021).

❖ ΑΝΤΙΔΡΑΣΕΙΣ ΠΟΥ ΜΕΣΟΛΑΒΟΥΝΤΑΙ ΑΠΟ ΑΝΤΙΣΩΜΑΤΑ (ΧΥΜΙΚΗ ΑΝΟΣΙΑ)

Η χυμική ανοσία (αντίδραση μεσολαβούμενη από αντισώματα) είναι ζωικής σημασίας για την πρόληψη των ιογενών λοιμώξεων.

Ένα υποσύνολο των αντισωμάτων, που παράγονται από τον οργανισμό κατά τη λοίμωξη με SARS-CoV-2 ονομάζονται εξουδετερωτικά αντισώματα. Αυτά τα αντισώματα συνδέονται με επιτόπους επιφανείας του ιού και μειώνουν τη μολυσματικότητα του. Επίσης, εμποδίζουν την είσοδο του ιού στο μολυσμένο κύτταρο. Έχει αναφερθεί ότι τα εξουδετερωτικά αντισώματα μειώνονται εντός 2-3 μηνών σε ασθενείς που έχουν αναρρώσει από την COVID-19. Σε μια μελέτη με 175 ασθενείς που ανάρρωσαν από ήπια νόσο COVID-19, τα ειδικά εξουδετερωτικά αντισώματα του SARS-CoV-2 ανιχνεύθηκαν στη φάση ανάρρωσης της λοίμωξης, 10 έως 15 ημέρες μετά την έναρξη της νόσου, και παρέμειναν έκτοτε (Huang et al., 2020, Alfego et al., 2021).

Φαίνεται ότι ανιχνεύονται συνήθως αντισώματα που δεσμεύουν την πρωτεΐνη N και την πρωτεΐνη-ακίδα του SARS-CoV-2. Η πρωτεΐνη-ακίδα είναι εξαιρετικά ανοσογόνος και ειδικά αντισώματα εναντίον της μπορούν να εξουδετερώσουν και να εμποδίσουν τις αλληλεπιδράσεις του ιού με τον υποδοχέα ACE2 (Zohar & Alter, 2020).

Έχει αποδειχθεί ότι η παραγωγή IgM αντισωμάτων έναντι του SARS-CoV-2 συμβαίνει και μεγιστοποιείται πριν από την παραγωγή αντισωμάτων IgG. Επιπλέον, η παραγωγή αντισωμάτων IgM αρχίζει να μειώνεται την τρίτη εβδομάδα της ασθένειας, ενώ η παραγωγή αντισωμάτων IgG επιμένει και διατηρείται (O'Connell & Aldhamen, 2020).

• ΚΑΤΑΡΡΑΚΤΗΣ ΚΥΤΤΑΡΟΚΙΝΩΝ

Οι ιογενείς λοιμώξεις μπορούν να οδηγήσουν σε υπερβολική δραστηριοποίηση του ανοσοποιητικού συστήματος προκαλώντας μαζική απελευθέρωση προφλεγμονωδών κυτταροκινών. Αυτό το φαινόμενο, γνωστό ως καταιγίδα κυτταροκινών, ξεκινά ως εντοπισμένη φλεγμονώδης απόκριση, αλλά στη συνέχεια επεκτείνεται συστηματικά, συμβάλλοντας στην περαιτέρω στρατολόγηση κυττάρων του ανοσοποιητικού σε ιστούς που έχουν μολυνθεί. Η ελεγχόμενη απελευθέρωση κυτταροκινών έχει βασικό ρόλο στην επίλυση της λοίμωξης, ενώ σε αντίθετη περίπτωση μπορεί να επέλθει ARDS και πολυοργανική ανεπάρκεια. Η καταιγίδα κυτταροκινών έχει παρατηρηθεί σε ασθενείς που έχουν μολυνθεί με SARS-CoV-2, οι οποίοι εκδηλώνουν σοβαρά συμπτώματα (Henderson et al., 2020).

Αρκετά ενεργοποιημένα κύτταρα του ανοσοποιητικού, όπως τα Β κύτταρα, τα Τ κύτταρα, τα NK κύτταρα, τα μακροφάγα, τα δενδριτικά κύτταρα, τα ουδετερόφιλα, τα μονοκύτταρα και τα ενδοθηλιακά κύτταρα, συμβάλλουν στην καταιγίδα των κυτταροκινών στη νόσο COVID-19. Σε σύγκριση με υγιείς μάρτυρες, ασθενείς που έχουν μολυνθεί με SARS-CoV-2 παρουσιάζουν αυξημένα επίπεδα προφλεγμονωδών κυτταροκινών και χημειοκινών στο πλάσμα, όπως IL-1β, IL-2, IL-6, IL-7, IL-10, παράγοντας νέκρωσης όγκου α (TNF-α), IFNγ, αυξητικός παράγοντας ινοβλαστών, αυξητικός παράγοντας που προέρχεται από αιμοπετάλια (Platelet-Derived Growth Factor, PDGF), αγγειακός ενδοθηλιακός αυξητικός παράγοντας, συνδέτης χημειοκίνης 2 (chemokine Ligand 2, CCL2), συνδέτης χημειοκίνης 3 (chemokine Ligand 3, CCL3), συνδέτης χημειοκίνης 4 (chemokine Ligand 3, CCL4), συνδέτης χημειοκίνης 8 (CCL8). Επιπλέον, έχει παρατηρηθεί μια σημαντική συσχέτιση

μεταξύ της σοβαρότητας της COVID-19 και των συγκεντρώσεων στον ορό των IL-1, IL-2, IL-6, IL-7, IL-10, TNF- α , G-CSF, CCL2, CCL3, IL-8. Είναι ενδιαφέρον ότι η αυξημένη συγκέντρωση TNF- α , IL-6 και IL-10 στον ορό ασθενών με σοβαρή νόσο COVID-19 σχετίζεται με μειωμένο αριθμό κυκλοφορούντων T κυττάρων που, επιπλέον, εμφανίζουν εξαντλημένο φαινότυπο, που χαρακτηρίζεται από υψηλή έκφραση PD-1 (Programmed Cell Death Protein 1) και Tim-3 (T Cell Immunoglobulin and Mucin Domain-Containing Protein 3). Έχει δειχθεί ότι ασθενείς με μέτρια και σοβαρή νόσο, παρουσιάζουν παρόμοια έκφραση φλεγμονωδών κυτταροκινών έως και 10 ημέρες μετά την έναρξη της νόσου, ενώ, σε μεταγενέστερα χρονικά σημεία, οι κυτταροκίνες TNF- α , IL-6, και τα επίπεδα IL-10 μειώνονται σταθερά σε ασθενείς με μέτρια νόσο και αντίθετα παραμένουν αυξημένα σε εκείνους με σοβαρή COVID-19. Μεταξύ των υπερεκφρασμένων κυτταροκινών και χημειοκινών σε ασθενείς με COVID-19, η IL-6 αντιπροσωπεύει έναν πολύτιμο βιοδείκτη λόγω της συσχέτισης του επιπέδου της στο πλάσμα τόσο με το ιικό φορτίο όσο και με πνευμονική βλάβη, σε ασθενείς που νόσησαν βαριά (Kim et al., 2020).

Τα BALFs (Bronchoalveolar lavage fluid) ασθενών με σοβαρή νόσο COVID-19 δείχνουν την υπερέκφραση των IL-6, IL-1 β και αρκετών χημειοκινών, όπως IL-8, κρίσιμης σημασίας για τη στρατολόγηση ουδετερόφιλων στον πνεύμονα, που φλεγμαίνει. Επιπλέον, οι πρωτεΐνες CCL2 και CCL8 έχουν βρεθεί αυξημένες σε μεταθανάτια δείγματα πνευμόνων από ασθενείς με COVID-19 σε σύγκριση με βιοψίες πνευμόνων από υγιείς μάρτυρες, επιβεβαιώνοντας έτσι το ρόλο της καταιγίδας κυτταροκινών στη σοβαρότητα της COVID-19 (A. K. G. Melo et al., 2021) .

Έτσι, με βάση τις τρέχουσες γνώσεις, η λεπτή και συντονισμένη δράση διαφορετικών διαλυτών διαμεσολαβητών, όπως ανπιφλεγμονώδεις κυτταροκίνες και χημειοκίνες, συμβάλλει τόσο στην εξέλιξη της λοίμωξης όσο και στις τοπικές και συστηματικές εκδηλώσεις της νόσου COVID-19 (Yang et al., 2021).

1.2 ΥΠΟΚΕΦΑΛΑΙΟ 2 - ΑΛΛΗΛΟΥΧΙΣΗ

1.2.1 ΑΛΛΗΛΟΥΧΙΣΗ RNA

Η αλληλούχιση RNA (RNA sequencing, RNA-SEQ) είναι μια τεχνική αλληλούχισης που χρησιμοποιεί αλληλούχιση επόμενης γενιάς (Next-generation sequencing, NGS), με σκοπό τον προσδιορισμό της παρουσίας και της ποσότητας του RNA σε ένα βιολογικό δείγμα σε μια δεδομένη χρονική στιγμή.

Η τυπική ροή της αλληλούχισης RNA αποτελείται από δύο μέρη, το πειραματικό μέρος και το υπολογιστικό μέρος. Το πειραματικό μέρος περιλαμβάνει την απομόνωση του RNA, τη μετατροπή του σε βιβλιοθήκες cDNA και την αλληλούχιση. Το υπολογιστικό μέρος περιλαμβάνει τη στοίχιση με ένα γονιδίωμα/μεταγράψωμα αναφοράς, τη διαφορική έκφραση και περαιτέρω αναλύσεις, όπως η λειτουργική ανάλυση (Cseke et al., 2003, Liang & Zeng, 2016).

Η γενική ροή εργασίας κατά την ανάλυση της γονιδιακής έκφρασης είναι:

1. Πειραματικό μέρος

- I. Απομόνωση RNA
- II. Προετοιμασία βιβλιοθήκης (συμπεριλαμβανομένου εμπλουτισμού mRNA)
- III. Αλληλούχιση επόμενης γενιάς
- IV. Έλεγχος ποιότητας

2. Βιοπληροφορική ανάλυση

- I. Επεξεργασία των αναγνώσεων (συμπεριλαμβανομένης της στοίχισης)
- II. Εκτίμηση των επιμέρους επιπέδων γονιδιακής έκφρασης
- III. Κανονικοποίηση

IV. Ταυτοποίηση γονιδίων με ανάλυση διαφορικής έκφρασης

• ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ

❖ ΑΠΟΜΟΝΩΣΗ RNA

Πριν από την αλληλούχιση του RNA, το RNA πρέπει πρώτα να απομονωθεί και να διαχωριστεί από το κυτταρικό του περιβάλλον, που αποτελείται κυρίως από πρωτεΐνες και DNA. Οι πιο διαδεδομένες μέθοδοι απομόνωσης RNA είναι οι μεμβράνες γέλης με βάση το πυρίτιο ή η απομόνωση με όξινη φαινόλη-χλωροφόρμιο. Στην πρώτη περίπτωση, το RNA συνδέεται αποκλειστικά με μεμβράνη πηκτής από πυρίτιο, ενώ τα υπόλοιπα κυτταρικά συστατικά ξεπλένονται. Οι μεμβράνες γέλης πυριτίου απαιτούν αιθανόλη για σύνδεση. Ο όγκος της αιθανόλης επηρεάζει ποια μετάγραφα συνδέονται στη μεμβράνη: περισσότερη αιθανόλη έχει ως αποτέλεσμα την κατακράτηση RNAs <200 bp, ενώ μικρότερος όγκος την απώλεια τους. Όταν χρησιμοποιείται απομόνωση φαινόλης-χλωροφορμίου, τα κυτταρικά συστατικά διαλύονται σε τρεις φάσεις: την οργανική φάση, την ενδιάμεση φάση και την υδατική φάση, στην οποία διατηρείται το RNA. Διαφορετικά άλατα οδηγούν σε διαφορετικές αποδόσεις κατακρήμνισης και έχουν ως αποτέλεσμα διαφορετικούς πληθυσμούς RNA. π.χ., το χλωρίδιο, ένα συνηθισμένο άλας, έχει αναφερθεί ότι έχει ως αποτέλεσμα την απώλεια tRNA, 5S rRNA, snRNA και άλλα RNAs <250-300 bp. Η ποσότητα της αποικοδόμησης του RNA ελέγχεται με πηκτική και τριχοειδική ηλεκτροφόρηση και χρησιμοποιείται για τον έλεγχο της ακεραιότητας του RNA στο δείγμα. Η ποιότητα του RNA και η συνολική ποσότητα αρχικού RNA λαμβάνονται υπόψη κατά τη διάρκεια των επόμενων σταδίων προετοιμασίας, προσδιορισμού αλληλουχίας και ανάλυσης (Devrimci-Ozguven et al., 2000, M et al., 2014)

❖ ΠΡΟΕΤΟΙΜΑΣΙΑ ΤΗΣ ΒΙΒΛΙΟΘΗΚΗΣ

1. Επιλογή/εξάλειψη RNA

Για την περαιτέρω ανάλυση, το απομονωμένο RNA μπορεί είτε να διατηρηθεί ως έχει, να φιλτραριστεί για να περιλαμβάνει μόνο mRNA με 3' πολυαδενυλιωμένες (πολύ-(A)) ουρές ή με εξάλειψη ριβοσωμικού RNA (rRNA), και/ ή φιλτράρεται για RNA που δεσμεύει συγκεκριμένες αλληλουχίες. Το RNA με πολύ-(A) ουρές αποτελείται κυρίως από ώριμες, επεξεργασμένες, κωδικοποιούσες αλληλουχίες. Η επιλογή πολυ-(A) πραγματοποιείται με ανάμιξη RNA με πολυ-(T) ολιγομερή συνδεδεμένα ομοιοπολικά σε ένα υπόστρωμα, τυπικά μαγνητικά σφαιρίδια (πολύ-(A) εμπλουτισμός, polyA enrichment). Η επιλογή πολυ-(A) έχει σημαντικούς περιορισμούς στην ανίχνευση συγκεκριμένου τύπου RNA. Πολλοί βιοτύποι RNA δεν είναι πολυαδενυλιωμένοι, συμπεριλαμβανομένων πολλών μη κωδικών μεταγραφών RNA και επομένως ενδέχεται να μην ανιχνευθούν μετά την επιλογή πολυ (A). Αυτοί οι περιορισμοί μπορούν να αποφευχθούν με την εξάλειψη του ριβοσωμικού RNA (rRNA depletion), δηλαδή με την αφαίρεση του rRNA που τυπικά αντιπροσωπεύει πάνω από το 90% του RNA σε ένα κύτταρο. Μικροί στόχοι RNA, όπως το miRNA, μπορούν να απομονωθούν περαιτέρω μέσω επιλογής μεγέθους με πηκτές αποκλεισμού, μαγνητικά σφαιρίδια ή εμπορικά kit.

2. Σύνθεση cDNA

Το RNA μεταγράφεται αντίστροφα σε cDNA επειδή το DNA είναι πιο σταθερό και επιτρέπει τον εμπλουτισμό του (με τη χρήση πολυμερασών DNA). Ο κατακερματισμός και η επιλογή μεγέθους εκτελούνται για τον καθαρισμό αλληλουχιών που είναι στο κατάλληλο μήκος για το μηχάνημα προσδιορισμού αλληλουχίας (αλληλουχητής - sequencer). Το RNA, το cDNA ή και τα δύο κατακερματίζονται με ένζυμα, υπερήχους ή νεφελοποιητές. Ο κατακερματισμός ακολουθείται από επιλογή μεγέθους, όπου είτε αφαιρούνται μικρές αλληλουχίες είτε επιλέγεται μια περιορισμένη σειρά μηκών

αλληλουχιών. Επειδή χάνονται μικρά RNA όπως τα miRNA, αυτά αναλύονται ανεξάρτητα (Head et al., 2014).

❖ ΑΛΛΗΛΟΥΧΙΣΗ

Σε όρους αλληλούχισης υψηλής απόδοσης, μια βιβλιοθήκη είναι μια (κατά προτίμηση τυχαία) συλλογή θραυσμάτων DNA που είναι έτοιμα για αλληλούχιση με ένα συγκεκριμένο πρωτόκολλο. Η πιο δημοφιλής πλατφόρμα είναι η HTS, που παρέχεται από την Illumina και βασίζεται στην αρχή «Sequencing by Synthesis», η οποία προήλθε από τα πρωτόκολλα αλληλούχισης Sanger που ξεκίνησαν τη δεκαετία του 1970. Αυτές οι μέθοδοι βασίζονται (αρχικά) σε ένζυμα που «διαβάζουν» το DNA συνεχώς, π.χ. DNA πολυμεράσες. Αυτά τα μόρια έχουν την ικανότητα να προσδιορίζουν την ακριβή σειρά με την οποία απαιτούνται να τοποθετηθούν συγκεκριμένα νουκλεοτίδια για να δημιουργηθεί ένα τέλειο συμπληρωματικό αντίγραφο ενός συγκεκριμένου κλώνου DNA. Η μέθοδος της Illumina βασίζεται ακριβώς σε αυτήν την ιδιότητα της DNA πολυμεράσης και απλώς διασφαλίζει ότι υπάρχει μια εγγραφή για τη βάση κάθε φορά που η πολυμεράση DNA προσθέτει ένα νέο νουκλεοτίδιο σε ένα εξελισσόμενο αντίγραφο DNA. Στις πλατφόρμες Illumina κάθε νουκλεοτίδιο επισημαίνεται με συγκεκριμένο φθοροφόρο έτσι ώστε η ενσωμάτωση των A, C, T, G να αποδίδει διακριτές εκπομπές. Αυτά τα σήματα καταγράφονται μέσω αυτόματων καμερών, που κάνουν χρήση εξελιγμένων μικροσκοπίων. Αυτό σημαίνει ότι προκειμένου να προσδιορισθεί η αλληλουχία των θραυσμάτων cDNA σε έναν προσδιοριστή ακολουθίας Illumina, τα θραύσματα πρέπει: 1) να είναι μονόκλωνα, 2) δεν μπορούν να υπερβαίνουν ένα ορισμένο μέγεθος, 3) πρέπει να ακινητοποιηθούν σε γυάλινη αντικειμενοφόρο πλάκα, και 4) οι εγγενείς ιδιότητες αλληλούχισης ενός δεδομένου θραύσματος (μήκος, περιεκτικότητα σε GC, ομοπολυμερή) θα επηρεάσουν την πιστότητα της DNA πολυμεράσης και άρα το απόλυτο σήμα. Ως εκ τούτου, η προετοιμασία της βιβλιοθήκης για τα πρωτόκολλα Illumina στοχεύει στη δημιουργία cDNA 150 έως 300 bp, το οποίο υβριδοποιείται σε γυάλινη πλάκα με μικρορευστονικούς θαλάμους (κυψελίδα ροής) (Stark et al., 2019).

✓ ΑΡΧΕΙΟ FASTQ

Οι ανεπεξέργαστες αναγνώσεις που προκύπτουν, από το πείραμα αλληλούχισης, πιο συχνά αποθηκεύονται ως αρχεία FASTQ. Η μορφή ενός FASTQ αρχείου, περιέχει την αλληλουχία κάθε μίας ανάγνωσης, που παράγεται από ένα πείραμα αλληλούχισης, μαζί με τα σκορ ποιότητας. Τα αρχεία FASTQ δεν είναι συμπιεσμένα και είναι αρκετά μεγάλα, επειδή περιέχουν τις ακόλουθες πληροφορίες για κάθε ανάγνωση (Cock et al., 2009):

1. @ που ακολουθείται από τον κωδικό (ID) της ανάγνωσης και πιθανόν πληροφορία για το πείραμα της αλληλούχισης
2. τις βάσεις που έχουν αλληλουχηθεί.
3. το σημείο +
4. σκορ ποιότητας για κάθε βάση της αλληλουχίας (ASCII - κωδικοποίηση)

❖ ΠΟΙΟΤΙΚΟΣ ΕΛΕΓΧΟΣ ΤΗΣ ΠΡΟΕΤΟΙΜΑΣΙΑΣ RNA

Το RNA είναι πολύ πιο ευαίσθητο στην αποικοδόμηση από το DNA. Η ποιότητα των μορίων RNA μπορεί να επηρεάσει έντονα τα αποτελέσματα του πειράματος RNA-SEQ. Παραδοσιακά, η ακεραιότητα του RNA εκτιμούνταν μέσω ηλεκτροφόρησης σε πηκτή με οπτική επισκόπηση των ριβοσωμικών ζωνών RNA. Άθικτο ευκαρυωτικό RNA θα πρέπει να αποδώσει καθαρές ζώνες rSR 28S και 18S. Κανονικά, η ζώνη 28S rRNA είναι περίπου διπλάσια από την ζώνη 18S rRNA ζώνη (αναλογία 2:1). Καθώς η ποιότητα του RNA μειώνεται, η αναλογία 2:1 μειώνεται και χαμηλού μοριακού βάρους RNA αρχίζει να συσσωρεύεται. Δεδομένου ότι η ανθρώπινη ερμηνεία των εικόνων γέλης είναι υποκειμενική και έχει αποδειχθεί ασυνεπής, ο Agilent ανέπτυξε έναν αλγόριθμο που επιτρέπει τον υπολογισμό ενός αριθμού ακεραιότητας RNA (RNA integrity number, RIN) από μια ψηφιακή αναπαράσταση της κατανομής μεγέθους των μορίων RNA (που μπορεί να ληφθεί από έναν Agilent Bioanalyzer). Ο αριθμός RIN βασίζεται σε ένα σύστημα αρίθμησης από το 1 έως το 10, με 1 το πιο υποβαθμισμένο ως προς την ποιότητα RNA και 10 το πιο άθικτο. Αυτή η προσέγγιση διευκολύνει την ερμηνεία και αναπαραγωγιμότητα, των αξιολογήσεων ποιότητας RNA, και παρέχει ένα

μέσο με το οποίο μπορούν τα δείγματα να συγκρίνονται με τυποποιημένο τρόπο (Schroeder et al., 2006).

• ΥΠΟΛΟΓΙΣΤΙΚΟ ΜΕΡΟΣ

❖ ΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΑΝΑΓΝΩΣΕΩΝ

Η επεξεργασία των αναγνώσεων περιλαμβάνει τον έλεγχο ποιότητας και τη χαρτογράφηση (W. V. Li & Li, 2018).

1. Έλεγχος ποιότητας

Αρχικά, είναι ωφέλιμο να πραγματοποιηθεί έλεγχος της ποιότητας των ανεπεξέργαστων αναγνώσεων. RNA με κακή ποιότητα θα χαρακτηρίζεται από την παρουσία ενός ή περισσότερων από τα ακόλουθα στοιχεία: διπλασιασμοί PCR (PCR duplicates), μόλυνση από προσαρμογέα (adapter contamination), αναγνώσεις rRNA & tRNA, μη χαρτογραφημένες αναγνώσεις (Williams et al., 2014).

2. Χαρτογράφηση

Προκειμένου να προσδιοριστούν τα μετάγραφα που υπάρχουν σε ένα συγκεκριμένο δείγμα και η γονιδιωματική προέλευση της αλληλουχίας, τα θραύσματα cDNA πρέπει να προσδιοριστούν. Η χαρτογράφηση των αλληλουχιών στο γονιδίωμα/μεταγράφημα αναφοράς συνιστά τη βασική διαδικασία μέσω της οποίας ποσοτικοποιούνται τα αποτελέσματα. Για καθεμία από τα εκατομμύρια των αναγνώσεων αλληλουχιών (sequence reads ή απλά reads) εντοπίζεται η θέση του γονιδιώματος ή του μεταγραφώματος από την οποία προέρχονται (το οποίο ονομάζεται γονιδίωμα/μεταγράφημα αναφοράς, reference genome/ reference transcriptome). Μετά τη χαρτογράφηση του συνόλου των αναγνώσεων, μπορούμε να γνωρίζουμε με ακρίβεια πόσες φορές “διαβάστηκε” κάθε νουκλεοτίδιο του υπό μελέτη γονιδιώματος. Η γενική πρόκληση της χαρτογράφησης εκατομμυρίων μικρών αναγνώσεων μετά από ένα πείραμα αλληλούχισης υψηλής απόδοσης, είναι η διαδικασία να

πραγματοποιηθεί σε εύλογο χρονικό διάστημα, με ακρίβεια. Τα περισσότερα εργαλεία για τη στοίχιση μικρών αναγνώσεων, χρησιμοποιούν αλγόριθμους, που ακολουθούν την προσέγγιση “seed-and-extend”. Αρχικά, ένα υποσύνολο της ανάγνωσης, χρησιμοποιείται ως ρίζα (“seed”), για την οποία το εργαλείο θα βρει το καλύτερο δυνατό ταίριασμα με ένα δείκτη, κατασκευασμένο με βάση το γονιδίωμα/ μεταγράψωμα αναφοράς (genome/ transcriptome index). Αυτή η ρίζα επεκτείνεται και προς τις δύο κατευθύνσεις, με ορισμένους περιορισμούς, μέχρι να στοιχηθεί όσο το δυνατόν περισσότερο γενετικό υλικό. Η τοπική στοίχιση, πραγματοποιείται συνήθως με τον αλγόριθμο Smith-Waterman algorithm (Smith and Waterman, 1981).

- ✓ Η παρουσία ενός γονιδιώματος/μεταγραφώματος αναφοράς καθώς και ένα αρχείο με το σχολιασμό των γονιδίων, είναι βασικό σε κάθε πείραμα RNA-SEQ. Οι γονιδιωματικές αλληλουχίες και ο γονιδιακός σχολιασμός παράγονται από κοινοπραξίες, όπως είναι οι κάτωθι: (mod)ENCODE, The Mouse Genome Project, The Berkeley Drosophila Genome Project. Αυτά τα αρχεία μπορούν να ανακτηθούν από την ιστοσελίδα της αντίστοιχης κοινοπραξίας. Οι UCSC και Ensembl προσπαθούν να οργανώσουν, να ενοποιήσουν και να παρέχουν πρόσβαση σε ένα εύρος γονιδιωμάτων και δεδομένων σχολιασμού. Οι αλληλουχίες αναφοράς συνήθως αποθηκεύονται σε απλό κείμενο FASTA, και μπορούν να συμπιεστούν με την γενική εντολή gzip (Conesa et al., 2016).

❖ ΕΚΤΙΜΗΣΗ ΤΩΝ ΕΠΙΜΕΡΟΥΣ ΕΠΙΠΕΔΩΝ ΓΟΝΙΔΙΑΚΗΣ ΕΚΦΡΑΣΗΣ

Παίρνοντας ως δεδομένες τις θέσεις των γονιδίων / μεταγράφων στο γονιδίωμα αναφοράς μπορούμε να υπολογίσουμε τον αριθμό των αναγνώσεων που επικαλύπτονται με κάθε γενετικό τόπο ξεχωριστά. Λαμβάνοντας υπόψη το μήκος της αντίστοιχης γονιδιωματικής περιοχής, (ή για την ακρίβεια των μεταγράφων mRNA που προκύπτουν από αυτήν), αλλά και το συνολικό αριθμό των αναγνώσεων που προέκυψαν από το πείραμα μπορούμε να καταλήξουμε σε μια αριθμητική τιμή που να αντιπροσωπεύει την

ποσότητα mRNA που υπήρχε στο αρχικό δείγμα από το συγκεκριμένο γενετικό τόπο (W. V. Li & Li, 2018).

❖ ΚΑΝΟΝΙΚΟΠΟΙΗΣΗ

Ο αριθμός των αναγνώσεων που επικαλύπτονται με ένα δεδομένο γονίδιο δεν μπορούν να αποδοθούν κατευθείαν ως απόλυτες τιμές γονιδιακής έκφρασης. Η τιμή, η οποία αποκτάται για ένα συγκεκριμένο γονίδιο σε ένα δείγμα βασίζεται στον αριθμό των αναγνώσεων που αντιστοιχούν σε ένα γονίδιο (ή μετάγραφο). Ωστόσο, υπάρχει μία σειρά παραμέτρων, που επηρεάζουν την αποτελεσματικότητα της ενίσχυσης και της αλληλούχισης των θραυσμάτων του DNA (πχ GC-bias). Σε ένα πείραμα RNA-SEQ, ο αριθμός των αλληλουχημένων αναγνώσεων στοιχισμένων με ένα γονίδιο, εξαρτάται εκτός από το επίπεδο έκφρασης του, από το μήκος του, το βάθος της αλληλούχισης και την έκφραση των λοιπών γονιδίων μέσα στο δείγμα. Έτσι ώστε να συγκρίνουμε την έκφραση ενός γονιδίου σε δύο συνθήκες, πρέπει να υπολογίσουμε το κλάσμα των αναγνώσεων, που αποδίδεται σε κάθε γονίδιο, σε σχέση με το συνολικό αριθμό των αναγνώσεων και την ποσότητα του RNA σε κάθε δείγμα. Ο σκοπός της κανονικοποίησης είναι να εξαλειφθούν οι συστηματικές επιδράσεις, που δεν οφείλονται σε βιολογικές διαφορές, καθώς και η εφαρμογή στατιστικού ελέγχου για τη ανάλυση διαφορικής έκφρασης (Risso et al., 2014).

❖ ΤΑΥΤΟΠΟΙΗΣΗ ΓΟΝΙΔΙΩΝ ΜΕ ΑΝΑΛΥΣΗ ΓΟΝΙΔΙΑΚΗΣ ΕΚΦΡΑΣΗΣ

Πολυάριθμες προσπάθειες έχουν γίνει με σκοπό την βελτιστοποίηση των στατιστικών ελέγχων, για να αποφασιστεί εάν η έκφραση ενός γονιδίου διαφέρει μεταξύ δύο καταστάσεων. Οι δύο βασικές λειτουργίες των εργαλείων διαφορικής έκφρασης είναι η εκτίμηση του μεγέθους της διαφορικής έκφρασης ανάμεσα σε δύο καταστάσεις με βάση τον αριθμό των αναγνώσεων στα δείγματα, λαμβάνοντας υπόψιν τις διαφορές στο βάθος της στοίχισης και τη μεταβλητότητα (variability), και δεύτερον η εκτίμηση της σημασίας της διαφοράς και η διόρθωση για πολλαπλούς ελέγχους (Yalamanchili et al., 2017) .

1.2.2 ΕΦΑΡΜΟΓΕΣ ΑΛΛΗΛΟΥΧΙΣΗΣ RNA

Οι εφαρμογές της αλληλούχισης RNA είναι οι κάτωθι:

1. Διαφορική έκφραση γονιδίων
2. Εντοπισμός πολυμορφισμών και έκφρασης διαφορετικών αλληλομόρφων, μετα-μεταγραφικές τροποποιήσεις, σύντηξη γονιδίων, μεταλλάξεις/SNPs
3. Ταυτοποίηση μικρών μορίων RNA
4. Χαρακτηρισμός προτύπων εναλλακτικού ματίσματος
5. Προσδιορισμός ορίων εξωνίου/εσωνίου και επαλήθευσης ή τροποποίησης των 5' και 3' ορίων των γονιδίων
6. Βιολογία Συστημάτων, ανάλυση μονοπατιών (pathway analysis), ανάλυση δικτύων (network analysis), λειτουργική ανάλυση διαφορικής έκφρασης, εμπλουτισμός γονιδιακών οντολογιών
7. Single-cell RNA-SEQ: μπορεί να αξιοποιηθεί για τη μελέτη της εμβρυογένεσης, τη μελέτη της διαφοροποίησης των βλαστοκυττάρων κτλ.(Picardi & Abeni, 2001, Differential Gene Expression (DGE) Analysis | Training-Modules, n.d.)

✓ ΔΙΑΦΟΡΙΚΗ ΕΚΦΡΑΣΗ ΓΟΝΙΔΙΩΝ

Ο στόχος της ανάλυσης διαφορικής έκφρασης είναι ο καθορισμός των γονιδίων που εκφράζονται σε διαφορετικά επίπεδα σε διάφορες συνθήκες. Η εύρεση αυτών των γονιδίων μπορεί να συντελέσει στον καθορισμό των διαδικασιών που επηρεάζονται από τη συνθήκη ενδιαφέροντος. Η ορθή ταυτοποίηση των διαφορικά εκφρασμένων γονιδίων ανάμεσα σε δύο ειδικές συνθήκες αποτελεί κομβικό σημείο για την κατανόηση της φαινοτυπικής ποικιλομορφίας (phenotypic variation). Ο υπολογισμός του βαθμού της μεταβολής της έκφρασης του ίδιου γονιδίου μεταξύ δύο διαφορετικών συνθηκών γίνεται με τη χρήση του λογαρίθμου του λόγου των πιμών έκφρασης στη συνθήκη μελέτης (test) προς τη συνθήκη ελέγχου (control). Η απόδοση των συνθηκών γίνεται από τον πειραματιστή και βασίζεται στο βιολογικό ερώτημα. Ο λογάριθμος του λόγου ονομάζεται ευρέως $\log_2\text{foldchange}$. Τιμές του λόγου >1 υποδεικνύουν μεγαλύτερη έκφραση στη

συνθήκη μελέτης και συνεπώς ενεργοποίηση του γονιδίου, ενώ τιμές <1 υποδεικνύουν καταστολή του (Αποθετήριο Κάλλιπος: Υπολογιστική Βιολογία, n.d.).

✓ ΓΟΝΙΔΙΑΚΕΣ ΟΝΤΟΛΟΓΙΕΣ

Η γονιδιακή οντολογία (gene ontology, GO) αποτελεί ένα από τα θεμελιώδη μέρη της αποτύπωσης του συνόλου της βιολογικής γνώσης σχετικά με τη γονιδιακή λειτουργία με ένα σαφή και οργανωμένο τρόπο.

Η βάση δεδομένων της GO έχει δημιουργήσει τρεις βασικές, ανεξάρτητες μεταξύ τους, κατηγορίες (οντολογίες) που περιγράφουν τα γονίδια όλων των μελετημένων οργανισμών στη βάση:

1. Της **βιολογικής διαδικασίας** (biological process) στην οποία συμμετέχουν οι πρωτεΐνες τους. Αφορά άρα στη λειτουργία του γονιδίου.
2. Της **μοριακής λειτουργίας** (molecular function) των πρωτεϊνών. Με ποιον τρόπο επιτελεί μοριακά τη λειτουργία της η πρωτεΐνη που αντιστοιχεί στο γονίδιο.
3. Του **κυτταρικού εντοπισμού** (cellular compartment) των πρωτεϊνών τους (Gene Ontology Resource, n.d.). Στο πλαίσιο της γονιδιακής οντολογίας, τα γονίδια αναπαρίστανται ως τμήματα-μονοπάτια στους κατευθυνόμενους ακυκλικούς γράφους που ορίζονται από την οντολογία.

Ο όρος εμπλουτισμός γονιδιακής οντολογίας (GO) είναι μια τεχνική για την ερμηνεία συνόλων γονιδίων που κάνουν χρήση του συστήματος ταξινόμησης Γονιδιακής Οντολογίας γονιδίων, στην οποία τα γονίδια αποδίδονται σε ένα σύνολο προκαθορισμένων ομάδων ανάλογα με τα λειτουργικά τους χαρακτηριστικά.

Υπάρχει μια ποικιλία μεθόδων για την εκτέλεση εμπλουτισμού χρησιμοποιώντας τις GO. Οι μέθοδοι διαφέρουν ανάλογα με τον τύπο της στατιστικής δοκιμής που εφαρμόζεται. Η πιο συνηθισμένη είναι η

υπεργεωμετρική κατανομή του Fisher. Άλλες μέθοδοι κάνουν χρήση της μπεύζιανής στατιστικής. Υπάρχει επίσης μεταβλητότητα στον τύπο διόρθωσης που εφαρμόζεται για πολλαπλές συγκρίσεις, με πιο συνηθισμένη τη διόρθωση Bonferroni (Αποθετήριο Κάλλιπος: Υπολογιστική Βιολογία, n.d., Consortium, 2008).

1.2.3 QUANTSEQ

Το QuantSeq, είναι ένα πρωτόκολλο, το οποίο παράγει δεδομένα έκφρασης γονιδίων με χαμηλό κόστος και χαμηλό θόρυβο. Το συγκεκριμένο πρωτόκολλο είναι επίσης γνωστό ως 3' Tag-Seq, TagSeq, 3' Tag RNA-SEQ, Digital RNA-SEQ. Σε αντίθεση με το παραδοσιακό RNA-SEQ, το οποίο οδηγεί στη δημιουργία βιβλιοθηκών αλληλούχισης, που αποτελούνται από ολόκληρα μετάγραφα, το QuantSeq παράγει μόνο ένα αρχικό μόριο βιβλιοθήκης ανά μετάγραφο, συμπληρωματικό στο 3' άκρο των αλληλουχιών (Ma et al., 2019).

1. Προετοιμασία του δείγματος

Το QuantSeq είναι ένα γρήγορο και ευέλικτο πρωτόκολλο, το οποίο παράγει βιβλιοθήκες NGS αλληλουχιών κοντά στο 3' άκρο πολυαδενυλιωμένων RNAs. Ο χρόνος που απαιτείται για αυτή τη διαδικασία είναι 4,5 ώρες, εκ των οποίων οι 2 είναι οι πραγματικές ώρες εκτέλεσης. Το κιτ απαιτεί μόνο 0.5–500 ng συνολικού RNA χωρίς να χρειάζεται poly(A) enrichment ή rRNA depletion. Λόγω της εστίασης στο 3' άκρο, το QuantSeq είναι κατάλληλο και για δείγματα σταθεροποιημένα σε φορμαλίνη/παραφίνη.

2. Αλληλούχιση

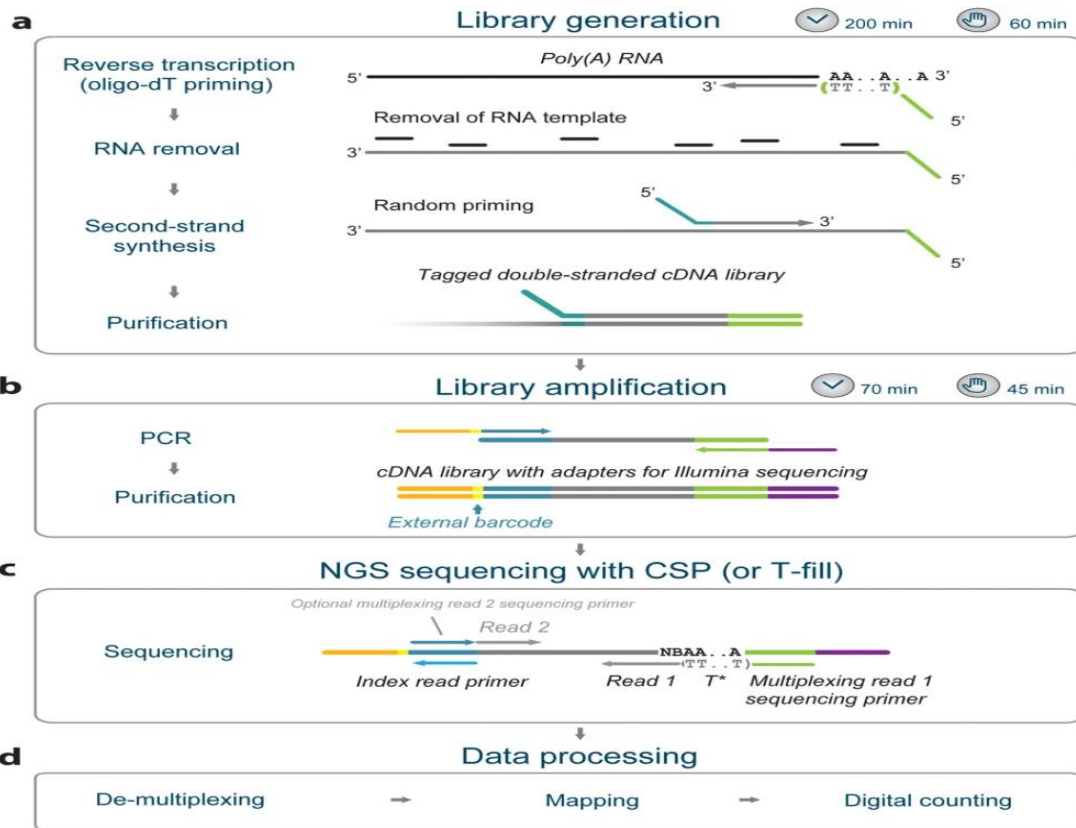
Το QuantSeq παράγει μόνο μία ανάγνωση ανά μετάγραφο, και ο αριθμός των αναγνώσεων που στοιχίζονται σε ένα δεδομένο γονίδιο είναι ανάλογος με την έκφραση του. Επίσης, λιγότερες αναγνώσεις απαιτούνται για τον καθορισμό πμών αναξιόπιστης έκφρασης, επιτρέποντας μεγαλύτερο βαθμό multiplexing (συνδυασμό δειγμάτων σε μία ανάλυση) και έτσι μείωση του κόστους.

3. Επεξεργασία των δεδομένων

Οι πιο πολλές αλληλουχίες προέρχονται από το τελευταίο εξώνιο και την 3' αμετάφραστη περιοχή, που περιέχει μόνο λίγες θέσεις ματίσματος. Αυτό έχει ως αποτέλεσμα τη δραματική μείωση του χρόνου στοίχισης. Το QuantSeq έχει υψηλή ειδικότητα κλώνου (>99.9%).

✓ Η ΡΟΗ ΕΡΓΑΣΙΑΣ ΤΟΥ QUANTSEQ

Η παραγωγή βιβλιοθήκης QuantSeq ξεκινά με την προετοιμασία των εκκινήτων (oligo-dT priming). Δεν απαιτείται προηγουμένως να έχει εκτελεστεί επιλογή/εξάλειψη RNA. Η σύνθεση του πρώτου κλώνου και η απομάκρυνση του RNA ακολουθείται από σύνθεση του συμπληρωματικού κλώνου με τυχαία εκκίνηση. Οι ειδικές συνδετικές αλληλουχίες Illumina ή IonTorrent (Illumina or IonTorrent-specific linker sequences) εισάγονται από τους εκκινήτες. Το DNA -με δύο κλώνους- που παράγεται, καθαρίζεται με μαγνητικά σφαιρίδια. Κατόπιν, η ενίσχυση της βιβλιοθήκης με PCR εισαγάγει τις πλήρεις αλληλουχίες, που απαιτούνται για τη δημιουργία των cluster. Οι βιβλιοθήκες Illumina μπορούν να κάνουν ανάμειξη δειγμάτων σε ένα πείραμα με μέχρι 96 εξωτερικούς κωδικούς και είναι συμβατές τόσο με αντιδραστήρια αλληλούχισης μονής ανάγνωσης (single-end) όσο και διπλής ανάγνωσης (paired-end). Το ιδανικό μέγεθος των αναγνώσεων είναι το μικρό (πχ, SR50 ή SR100). Οι βιβλιοθήκες IonTorrent μπορούν να συνδυαστούν με τη χρήση 24 barcodes (Εικόνα 4).



Εικόνα 4. Η ροή εργασίας του QuantSeq

1.2.4 ΣΥΓΚΡΙΣΗ RNA-SEQ – QUANTSEQ

Το QuantSeq είναι μία αξιόπιστη και απλή μέθοδος RNA-SEQ. Αυξάνει την ακρίβεια στις μετρήσεις της γονιδιακής έκφρασης, καθώς παράγεται μόνο μία ανάγνωση ανά μετάγραφο. Σε μικρότερα βάθη, αυτή η εστίαση στο 3' άκρο οδηγεί σε υψηλότερη σταθερότητα στις μετρήσεις της γονιδιακής έκφρασης. Το QuantSeq είναι ιδανικό για την αύξηση του βαθμού της ανάμειξης δειγμάτων σε ένα πείραμα (multiplexing) στα NGS πειράματα γονιδιακής έκφρασης και είναι η μέθοδος εκλογής για τον ακριβή προσδιορισμό της γονιδιακής έκφρασης με χαμηλότερο κόστος. Το QuantSeq έχει παρόμοια αναπαραγωγιμότητα με το total RNA-SEQ. Το παραδοσιακό RNA-SEQ ανιχνεύει πιο λίγες αναγνώσεις όταν μειώνεται το βάθος της στοίχισης. Το παραδοσιακό RNA-SEQ επηρεάζεται από το μήκος του μεταγράφου, ενώ στο QuantSeq το ένα μετάγραφο αντιστοιχεί σε μία ανάγνωση. Το QuantSeq είναι κατάλληλο όταν ο ερευνητής ενδιαφέρεται για την διαφορική έκφραση και όχι για τον προσδιορισμό των ορίων εξωνίου/εσωνίου και της αλληλουχίας του μεταγράφου (Schwartz et al., 2016).

Τα πλεονεκτήματα του QuantSeq είναι τα κάτωθι:

1. Στο QuantSeq η ποιότητα του RNA των δειγμάτων δεν είναι τόσο σημαντική όσο σε άλλες εφαρμογές του RNA seq. Με το QuantSeq, χαμηλή σχετικά ποιότητα RNA μπορεί να επιφέρει καλά αποτελέσματα. Γενικά, οι απαιτήσεις ως προς την ποιότητα και την ποσότητα των δειγμάτων είναι μικρές (100pg).
2. Πιο ευέλικτο και πιο οικονομικό πρωτόκολλο προετοιμασίας της βιβλιοθήκης
3. Διατηρεί την πληροφορία προέλευσης από κάθε αλυσίδα του DNA (strand specific)
4. Μπορούμε να αποκτήσουμε ακριβείς πληροφορίες για την 3' αμετάφραστη περιοχή (REV kit)
5. Η αλληλούχιση μονής κατεύθυνσης είναι επαρκής για την ποσοτικοποίηση της έκφρασης των γονιδίων

Τα μειονεκτήματα του QuantSeq είναι τα εξής:

1. Τα δεδομένα δεν περιέχουν πληροφορίες για το μάπισμα των μεταγράφων
2. Απαιτείται η χρήση ενός γονιδιώματος αναφοράς με καλό σχολιασμό (και γνωστές αμετάφραστες περιοχές)
3. Είναι δυνατόν να εφαρμοστεί μόνο σε δείγματα από ευκαρυωτικούς οργανισμούς (απαιτείται πολύ-(A) εμπλουτισμός)
4. Το πρωτόκολλο του QuantSeq είναι πιο ευαίσθητο σε χημικούς μολυντές

1.3 ΥΠΟΚΕΦΑΛΑΙΟ 3 - ΕΡΕΥΝΗΤΙΚΟ ΕΡΩΤΗΜΑ – ΣΚΟΠΟΣ ΤΗΣ ΕΡΓΑΣΙΑΣ

Το ερευνητικό ερώτημα που τέθηκε ήταν ότι το παρατεταμένο θετικό RT-PCR SARS-CoV-2 test σε ασθενείς με νόσο COVID-19 συσχετίζεται με διαφορά στην ανοσολογική απόκριση κατά την αρχική φάση της λοίμωξης. Για το σκοπό αυτό, ρινοφαρυγγικά δείγματα από 8 ασθενείς με νόσο COVID-19 επεξεργάστηκαν και αναλύθηκαν. Σε 4 από αυτούς τους ασθενείς, το RT-PCR SARS-CoV-2 test εξακολουθούσε να είναι θετικό ένα μήνα μετά την πρωτοδιάγνωση της λοίμωξης, ενώ στους άλλους 4 το δεύτερο RT-PCR SARS-CoV-2 test ήταν αρνητικό μία εβδομάδα μετά την αρχική θετικοποίηση του test. Σκοπός της παρούσας εργασίας ήταν η αλληλούχιση του RNA των άνωθι δειγμάτων με τη χρήση του πρωτοκόλλου QuantSeq και η ανάπτυξη βιοπληροφορικής ροής (pipeline) για τη βιοπληροφορική ανάλυση των δεδομένων που προέκυψαν, με στόχο τη μελέτη διαφορών στην ανοσολογική απόκριση στις δύο ομάδες ασθενών. Ειδικότερα, η διαδικασία που ακολουθήθηκε ήταν η εξής: RNA απομόνωση, RNA αλληλούχιση (QuantSeq), βιοπληροφορική ανάλυση των δειγμάτων (trimming με το εργαλείο BBduk, στοίχιση με τα εργαλεία STAR/SALMON, ανάλυση διαφορικής έκφρασης με τα εργαλεία DESeq2/edgeR, λειτουργική ανάλυση με τα εργαλεία DAVID/g:Profiler).

ΚΕΦΑΛΑΙΟ 2 - ΜΕΘΟΔΟΣ

2.1 ΔΕΙΓΜΑΤΑ

Στην εργασία χρησιμοποιήθηκαν 8 δείγματα, τα οποία συνίστανται σε αποθηκευμένο πλήρως ανώνυμο RNA από ασθενείς με COVID-19. Ειδικότερα, τα δείγματα, προέρχονταν από το ρινοφάρυγγα ασθενών με COVID-19. Τα δείγματα κωδικοποιήθηκαν ως εξής: 0728, 0729, 0730, 0732, 0733, 0735, 0737, 0739.

Πραγματοποιήθηκε η αλληλούχιση των με το πρωτόκολλο της QuantSeq, σε αλληλουχητή Illumina. Τα εξής τέσσερα δείγματα (0728, 0729, 0730, 0735) προέρχονται από ασθενείς, που είχαν θετικό PCR test για SARS-CoV-2, ένα μήνα μετά την αρχική θετικοποίηση του τεστ, ενώ τα υπόλοιπα τέσσερα δείγματα (0732, 0733, 0737, 0739), προέρχονταν από ασθενείς που είχαν αρνητικό δεύτερο τεστ. Τα δείγματα κατηγοριοποιήθηκαν σε δύο ομάδες (groups), την ομάδα A, στην οποία εντάχθηκαν τα δείγματα 0728, 0729, 0730, 0735 - που προέρχονταν από ασθενείς με παρατεταμένο θετικό test-, ενώ στην ομάδα B εντάχθηκαν τα δείγματα 0732, 0733, 0737, 0739 - που προέρχονταν από ασθενείς με αρνητικό δεύτερο test. Η ηλικία των συμμετεχόντων είχε εύρος 28-79 έτη. Η διάμεση ηλικία ήταν 39 έτη και η μέση ηλικία 44 έτη. Η διάμεση ηλικία των ασθενών που ανήκαν στην ομάδα A ήταν 33 έτη, ενώ η μέση ήταν 37. Η διάμεση ηλικία των ασθενών που ανήκαν στην ομάδα B ήταν 51 έτη, ενώ η μέση ήταν 48 έτη. Αναφορικά με το φύλο των συμμετεχόντων, 5/8 ήταν άνδρες, και 3/8 γυναίκες. Στην ομάδα A, ανήκαν 3 άνδρες και 1 γυναίκα. Στην ομάδα B, ανήκαν 2 άνδρες και 2 γυναίκες. Σχετικά με συνοσηρότητες, ένας ασθενής από την ομάδα A και ένας ασθενής από την ομάδα B ανέφεραν νόσο του καρδιαγγειακού συστήματος και χρόνια αναπνευστική πνευμονοπάθεια αντίστοιχα. Ένας ασθενής από την ομάδα A, χρειάστηκε νοσηλεία στο νοσοκομείο. Τέσσερις συνολικά ασθενείς εμφάνισαν πυρετό, δύο από την ομάδα A και δύο από την ομάδα B. Λοιπές κλινικές πληροφορίες δυστυχώς δεν ήταν διαθέσιμες για τα δείγματα (Πίνακας 1).

Πίνακας 1. Πληροφορίες για τα δείγματα που χρησιμοποιήθηκαν (κωδικός δείγματος, αλληλουχητής, μέθοδος, αναγνώσεις, ομάδα, φύλο, ηλικία)

Κωδικός Δείγματος	Αλληλουχητής	Μέθοδος	Αναγνώσεις	Ομάδα	Φύλο	Ηλικία
NGS-0728	ILLUMINA	3'QuantSeq	9818621	A	Άρρεν	32
NGS-0729	ILLUMINA	3'QuantSeq	11843440	A	Άρρεν	54
NGS-0730	ILLUMINA	3'QuantSeq	10341789	A	Άρρεν	33
NGS-0732	ILLUMINA	3'QuantSeq	5950275	B	Θήλυ	79
NGS-0733	ILLUMINA	3'QuantSeq	8268443	B	Άρρεν	45
NGS-0735	ILLUMINA	3'QuantSeq	9728524	A	Θήλυ	28
NGS-0737	ILLUMINA	3'QuantSeq	10839535	B	Άρρεν	50
NGS-0739	ILLUMINA	3'QuantSeq	12846800	B	Θήλυ	29

2.2 ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ

Αρχικά, πραγματοποιήθηκε η απόμονωση του RNA από τα 8 ρινοφαρυγγικά δείγματα. Κατόπιν, ακολούθησε, η προετοιμασία της βιβλιοθήκης QuantSeq. Η δημιουργία της βιβλιοθήκης ξεκίνησε με την προετοιμασία των εκκινητών (oligo dT priming). Ο εκκινητής αυτός περιείχε ήδη συνδετικές αλληλουχίες συμβατές με Illumina. Μετά τη σύνθεση του πρώτου κλώνου, το RNA αφαιρέθηκε και ξεκίνησε η σύνθεση του συμπληρωματικού κλώνου DNA με τη χρήση της DNA πολυμεράσης. Ο τυχαίος εκκινητής περιείχε επίσης συνδετικές αλληλουχίες συμβατές με Illumina. Η σύνθεση του cDNA ακολουθήθηκε από ένα βήμα καθαρισμού με βάση ένα μαγνητικό σφαιρίδιο (magnetic bead). Η βιβλιοθήκη, στη συνέχεια, ενισχύθηκε και έγινε εισαγωγή των ακολουθιών που απαιτούνται για τη δημιουργία cluster (cluster

generation). Τα εξωτερικά barcodes περιλαμβάνονται στο kit QuantSeq και εισήχθησαν κατά τη διάρκεια του σταδίου ενίσχυσης PCR. Η ποσοτικοποίηση της βιβλιοθήκης πραγματοποιήθηκε από τον αλληλουχητή (sequencer).

Τα βήματα, που ακολουθήθηκαν, πιο αναλυτικά, είναι τα κάτωθι:

1. Δημιουργία της βιβλιοθήκης (αντίστροφη μεταγραφή, απομάκρυνση RNA, σύνθεση δεύτερου κλώνου, καθαρισμός)

• Σύνθεση του πρώτου κλώνου του cDNA- Αντίστροφη μεταγραφή

Ένας εκκινητής oligodT που περιέχει αλληλουχία συμβατή με Illumina, στο 5' άκρο του υβριδοποιείται στο RNA και πραγματοποιείται αντίστροφη μεταγραφή.

• Απομάκρυνση του RNA

Κατά τη διάρκεια αυτού του βήματος η μήτρα του RNA (RNA template) αποικοδομήθηκε

• Σύνθεση του δεύτερου κλώνου

Κατά τη διάρκεια αυτού του βήματος, η βιβλιοθήκη μετατράπηκε σε dsDNA. Η σύνθεση του δεύτερου κλώνου ξεκίνησε με ένα τυχαίο εκκινητή (random primer), που περιείχε μία συνδετική αλληλουχία συμβατή με την Illumina στο 5' ακρο.

• Καθαρισμός

Η βιβλιοθήκη καθαρίστηκε με τη χρήση μαγνητικών σφαιριδίων με σκοπό την απομάκρυνση όλων των συστατικών της αντίδρασης.

2. Ενίσχυση της βιβλιοθήκης

• Προετοιμασία

Πραγματοποιήθηκε PCR. Η βιβλιοθήκη ενισχύθηκε με σκοπό την προσθήκη πλήρων αλληλουχιών adapters για τη δημιουργία των clusters και για την παραγωγή επαρκούς υλικού για έλεγχο ποιότητας και αλληλούχιση.

- Καθαρισμός

Η ολοκληρωμένη βιβλιοθήκη καθαρίστηκε από τα στοιχεία της PCR.

3. Αλληλούχιση

Πραγματοποιήθηκε FWD αλληλούχιση μονής ανάγνωσης (single-end).

4. Έλεγχος ποιότητας RNA

Η ακεραιότητα των δειγμάτων RNA εκτιμήθηκε με τη χρήση μιας δοκιμασίας μικρορευστοποίησης, τη σειρά RNA6000 για το 2100 Bioanalyzer (Agilent Technologies, Inc.), αν και η ποιότητα του RNA μπορεί επίσης να εκτιμηθεί με μετουσίωση της ηλεκτροφόρησης της γέλης αγαρόζης εάν μια τέτοια συσκευή δεν είναι διαθέσιμη. Καθώς το QuantSeq στοχεύει συγκεκριμένα στο 3' άκρο των μεταγραφών, ακόμη και RNA με χαμηλότερο RIN είναι κατάλληλα ως υλικό εισόδου.

2.3 ΥΠΟΛΟΓΙΣΤΙΚΟ ΜΕΡΟΣ

Κατόπιν της εκτέλεσης του πειραματικού μέρους, πραγματοποιήθηκε βιοπληροφορική ανάλυση αρχικά σε περιβάλλον Unix -FastQC, BBDuk, STAR, Salmon-, και έπειτα σε περιβάλλον R -DESeq2, edgeR-. Τέλος, έγινε χρήση διαδικτυακών εργαλείων (DAVID, g:Profiler, CIBERSORT-X, xCell).

2.3.1 UNIX

Το Unix ή UNIX είναι λειτουργικό σύστημα Ηλεκτρονικών Υπολογιστών, το οποίο αναπτύχθηκε κατά τις δεκαετίες του 1960 και του 1970 από ομάδα εργαζομένων των εργαστηρίων Μπελ. Σήμερα τα συστήματα Unix έχουν χωριστεί σε πολλούς κλάδους και αναπτύσσονται τόσο από την AT&T όσο και

από άλλους εμπορικούς παράγοντες, όπως και από αρκετούς μη κερδοσκοπικούς οργανισμούς, όπως το πρόγραμμα GNU. Το Unix σχεδιάστηκε για να μεταφέρεται εύκολα σε άλλες πλατφόρμες και να υποστηρίζει πολλαπλές ταυτόχρονες εργασίες παράλληλα με την ταυτόχρονη χρήση του από πολλούς χρήστες, με διάταξη χρονομερισμού. Τα συστήματα Unix χαρακτηρίζονται συνήθως από τις εξής ιδιότητες: χρήση απλού κειμένου για την αποθήκευση των δεδομένων, ιεραρχικό σύστημα αρχείων, χρήση ενός μεγάλου αριθμού εργαλείων, μικρές εφαρμογές που μπορούν να συνδυαστούν με ένα διερμηνέα γραμμής εντολών με χρήση σωληνώσεων (pipes).

2.3.2 BASH

Η Bash είναι κέλυφος (shell) και γλώσσα εντολών του Unix που γράφτηκε από τον Brian Fox για το GNU Project. Κυκλοφόρησε για πρώτη φορά το 1989 και χρησιμοποιήθηκε ως το προεπιλεγμένο κέλυφος σύνδεσης για τις περισσότερες διανομές Linux. Το όνομα είναι αρκτικόλεξο για το «Bourne-Again SHell». Η Bash είναι ευρέως διαθέσιμη σε διάφορα λειτουργικά συστήματα και είναι ένας προεπιλεγμένος διερμηνέας εντολών στα περισσότερα συστήματα GNU/Linux. Η Bash είναι ένας διερμηνέας γλώσσας εντολών που συνήθως εκτελείται σε ένα παράθυρο κειμένου όπου ο χρήστης πληκτρολογεί εντολές που προκαλούν ενέργειες. Η Bash είναι επίσης γλώσσα προγραμματισμού και μπορεί να διαβάσει και να εκτελέσει εντολές από ένα αρχείο. Οι λέξεις-κλειδιά, η σύνταξη, οι δυναμικά καθορισμένες μεταβλητές και άλλα βασικά χαρακτηριστικά της γλώσσας συνοδεύονται από το sh. Η Bash υποστηρίζει μεταβλητές, συναρτήσεις και έχει δομές ροής ελέγχου, όπως δηλώσεις υπό όρους και βρόχους (Bash - GNU Project - Free Software Foundation, n.d.).

2.3.3 ΓΛΩΣΣΑ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ R

Η R είναι γλώσσα προγραμματισμού και ελεύθερο περιβάλλον λογισμικού για στατιστικούς υπολογισμούς και γραφικά που υποστηρίζεται από την R Core

Team και το R Foundation for Statistical Computing. Χρησιμοποιείται ευρέως για την ανάπτυξη λογισμικών στατιστικής, μηχανικής μάθησης και ανάλυσης δεδομένων. Από τον Αύγουστο του 2021, η γλώσσα προγραμματισμού R κατατάσσεται στην 14η θέση στον δείκτη TIOBE, ένα μέτρο δημοτικότητας των γλωσσών προγραμματισμού.

Το επίσημο περιβάλλον λογισμικού R είναι ένα πακέτο GNU. Γράφτηκε κυρίως σε C, Fortran και R και διατίθεται ελεύθερα υπό την GNU General Public License. Παρέχονται προεγκατεστημένα εκτελέσιμα για διάφορα λειτουργικά συστήματα. Διαθέτει διεπαφή γραμμής εντολών, αλλά είναι διαθέσιμα πολλά γραφικά περιβάλλοντα χρήστη τρίτων κατασκευαστών, όπως το RStudio, ένα ολοκληρωμένο περιβάλλον ανάπτυξης. Οι χρήστες έχουν συνήθως πρόσβαση μέσω διερμηνέα γραμμής εντολών.

Οι δυνατότητες της R επεκτείνονται μέσω πακέτων που δημιουργούνται από χρήστες, τα οποία επιτρέπουν εξειδικευμένες στατιστικές αναλύσεις, γραφικές αναπαραστάσεις, και δίνουν πρόσβαση σε διάφορα εργαλεία (RMarkdown, knitr, Sweave) κλπ. Ο μεγάλος αριθμός των διαθέσιμων πακέτων και η ευκολία στην εγκατάσταση και χρήση τους, έχει αναφερθεί ως ένας σημαντικός παράγοντας στην προώθηση της ευρείας υιοθέτησης της γλώσσας στην επιστήμη των δεδομένων.

Με τη βασική εγκατάσταση περιλαμβάνεται ένα βασικό σύνολο πακέτων, ενώ υπάρχουν διαθέσιμα περισσότερα από 15.000 πρόσθετα πακέτα. Είναι διαθέσιμα στο Συνολικό δίκτυο αρχειοθέτησης R (CRAN), Bioconductor, Omegahat, GitHub και άλλα αποθετήρια.

2.3.4 BIOCONDUCTOR

Το Bioconductor παρέχει εργαλεία για τη διερεύνηση και την ανάλυση γονιδιωματικών δεδομένων υψηλής απόδοσης. Το Bioconductor χρησιμοποιεί τη στατιστική γλώσσα προγραμματισμού R. Είναι ανοιχτού κώδικα και αναπτύσσεται συνεχώς. Έχει δύο κυκλοφορίες κάθε χρόνο και μια ενεργή

κοινότητα χρηστών. Το Bioconductor είναι επίσης διαθέσιμο στο Docker. Ο στόχος του Bioconductor είναι να προωθήσει τη στατιστική ανάλυση και κατανόηση των τρεχουσών και αναδυόμενων βιολογικών δοκιμασιών υψηλής απόδοσης. Το Bioconductor βασίζεται σε πακέτα γραμμένα κυρίως στη γλώσσα προγραμματισμού R. Το Bioconductor δεσμεύεται για ανοικτού κώδικα, συνεργατική, κατανεμημένη ανάπτυξη λογισμικού και αναπαραγωγίμη έρευνα (Bioconductor - About, n.d.).

2.3.5 FASTQC

Το FastQC είναι ένα πρόγραμμα που έχει σχεδιαστεί για να εντοπίζει πιθανά προβλήματα σε σύνολα γονιδιωματικών δεδομένων υψηλής απόδοσης. Εκτελεί ένα σύνολο αναλύσεων σε ένα ή περισσότερα ανεπεξέργαστα αρχεία, που περιέχουν αλληλουχίες, σε μορφή fastq ή bam και παράγει μια έκθεση που συνοψίζει τα αποτελέσματα (Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data, n.d.).

2.3.6 BBDUK

Το BBDuk είναι ένα εργαλείο, το οποίο αναπτύχθηκε για να συνδυάσει τις λειτουργίες του trimming και φιλτραρίσματος των δεδομένων σε ένα μοναδικό εργαλείο. Το “Duk” σημαίνει Decontamination Using Kmers. Το BBDuk πραγματοποιεί υψηλής ποιότητας trimming και φιλτράρισμα, trimming των adapters (προσαρμογέας), φιλτράρισμα των GC, φιλτράρισμα μήκους, φιλτράρισμα εντροπίας, μετατροπή του format, και πολλές άλλες λειτουργίες σε ένα μόνο πέρασμα του αλγορίθμου (*BBDuk Guide - DOE Joint Genome Institute*, n.d.).

2.3.7 STAR

Το STAR (Spliced Transcripts Alignment to a Reference) είναι ένα γρήγορο εργαλείο στοίχισης δεδομένων RNA-SEQ, που λαμβάνει υπόψιν του τα εναλλακτικά μετάγραφα, που προκύπτουν μετά από το μάπισμα. Διαφορετικά

μέρη μίας ανάγνωσης μπορούν να στοιχηθούν σε διαφορετικές γενωμικές θέσεις, που αντιστοιχούν στη θέση του ματίσματος. Ο δείκτης του γονιδιώματος περιλαμβάνει γνωστές θέσεις ματίσματος από καλά σχολιασμένα γονίδια (annotated genes). Το STAR εκτελεί τοπική στοίχιση, και αυτόματη εξάλειψη των άκρων με υψηλή αναντιστοιχία (Dobin et al., 2013).

Η βασική ροή του STAR αποτελείται από δύο βήματα:

1. Παραγωγή του αρχείου δείκτης του γονιδιώματος (genome index)

Σε αυτό το βήμα ο χρήστης παρέχει την αλληλουχία του γονιδιώματος αναφοράς (FASTA αρχείο) και το σχολιασμό του (GTF αρχείο). Τα αρχεία GTF ορίζουν γονιδιωματικές περιοχές που αντιστοιχούν σε διαφορετικούς τύπους γενωμικής πληροφορίας, π.χ. γονίδια, μεταγραφές, εξόνια ή αμετάφραστες περιοχές. Έτσι το STAR παράγει ένα αρχείο το οποίο περιέχει τους δείκτες του γονιδιώματος και το οποίο θα χρησιμοποιηθεί στο δεύτερο βήμα, στη χαρτογράφηση.

2. Στοίχιση των αναγνώσεων στο γονιδίωμα

Σε αυτό το βήμα, ο χρήστης παρέχει το αρχείο που δημιουργήθηκε στο πρώτο βήμα, καθώς και τις αναγνώσεις του RNA-SEQ (αλληλουχίες) σε μορφή FASTA ή FASTQ. Το STAR στοιχίζει τις αναγνώσεις στο γονιδίωμα και παρέχει μία σειρά αρχείων, όπως για παράδειγμα τις στοιχίσεις (SAM, BAM), περίληψη των στατιστικών της στοίχισης, τις μη στοιχισμένες αναγνώσεις.

- Τα αρχεία που δημιουργεί το STAR είναι τα κάτωθι (Dobin, 2019):

1. **Αρχεία log:** περιγράφουν τη διαδικασία που ακολουθήθηκε
2. **Αρχεία SAM:** Aligned.out.sam - στοιχίσεις σε τυπική μορφή SAM

3. Ταξινομημένα και μη, με βάση τη συντεταγμένη, αρχεία **BAM** (unsorted και sorted BAM files)
4. Θέσεις ματίσματος
5. **ReadsPerGene.out.tab**: Με τη χρήση της παραμέτρου `--quantMode GeneCounts` το STAR παράγει ένα επιπλέον αρχείο με τις αναγνώσεις ανά γονίδιο, που προέκυψαν κατά τη χαρτογράφηση. Η δομή του συγκεκριμένου αρχείου είναι η ακόλουθη:
 - 1η στήλη: ο κωδικός του γονιδίου (gene ID column),
 - 2η στήλη: ο αριθμός των μεταγράφων για δεδομένα από RNA-SEQ οποιαδήποτε κατεύθυνσης,
 - 3η στήλη: ο αριθμός των μεταγράφων για δεδομένα RNA-SEQ μονής ανάγνωσης στοιχισμένα με τον πρώτο κλώνο,
 - 4η στήλη: αριθμός των μεταγράφων για δεδομένα RNA-SEQ μονής ανάγνωσης, στοιχισμένα με τον δεύτερο κλώνο

2.3.8 SALMON

Το Salmon είναι ένα εργαλείο, το οποίο στοχεύει στο γρήγορο ποσοτικό προσδιορισμό των μεταγράφων σε δεδομένα αλληλούχισης RNA. Απαιτείται ένα μεταγράψωμα αναφοράς για ποσοτικοποίηση. Το μόνο που χρειάζεται για να εκτελεστεί το Salmon είναι ένα αρχείο FASTA που περιέχει το μεταγράψωμα αναφοράς και ένα σύνολο αρχείων FASTA/FASTQ που περιέχει τις αναγνώσεις. Η στοίχιση εκτελείται σε δύο φάσεις: παραγωγή του δείκτη του γονιδιώματος (genome index) και ποσοτικοποίηση (quantification) (Patro et al., 2017).

2.3.9 DESEQ2

Το πακέτο DESeq2 είναι σχεδιασμένο για κανονικοποίηση, οπτικοποίηση, και διαφορική έκφραση πολυδιάστατων ποσοτικών δεδομένων. Χρησιμοποιεί τεχνικές μπεύζιανής στατιστικής, με σκοπό να εκτιμήσει το λογαριθμικό λόγο

αλλαγής (log fold change) και τη διασπορά για ποσοτικά δεδομένα αλληλούχισης. Λειτουργεί σε περιβάλλον γλώσσας προγραμματισμού R. Οι συγγραφείς του πακέτου είναι οι κάτωθι : Michael Love, Wolfgang Huber, Simon Anders (Love et al., 2014).

Οι κύριες λειτουργίες του είναι (Michael et al., 2021):

1. **DESeqDataSet**: σχηματίζει το σύνολο δεδομένων
2. **DESeq**: εκτελεί την ανάλυση διαφορικής έκφρασης βασισμένο στην αρνητική διωνυμική κατανομή (γάμμα - Poisson). Αυτή η λειτουργία πραγματοποιεί ανάλυση μέσω των κάτωθι βημάτων: 1. εκτίμηση των παραγόντων μεγέθους: estimateSizeFactors 2. εκτίμηση της διασποράς: estimateDispersions 3. Χρήση αρνητικής διωνυμικής κατανομής GLM fitting και στατιστική Wald: nbinomWaldTest
3. **results**: δημιουργεί έναν πίνακα αποτελεσμάτων, που προκύπτει από την ανάλυση DESeq. Ειδικότερα, αυτός ο πίνακας, που περιέχει μεταξύ άλλων, για κάθε γονίδιο τα κάτωθι: το λογαριθμικό λόγο αλλαγής ($\log_2\text{foldchange}$), το τυπικό σφάλμα, την τιμή p (p-value), τη διορθωμένη τιμή p (adjusted p-value). Η λειτουργία coef είναι σχεδιασμένη για προχωρημένους χρήστες που επιθυμούν να εξετάσουν όλους τους συμπαράγοντες ενός μοντέλου την ίδια στιγμή.
4. **LfcShrink**
5. **Vst**
6. **Γραφήματα**: για παράδειγμα plotPCA, plotMA, plotCounts

2.3.10 EDGER

Το edgeR είναι ένα πακέτο λογισμικού Bioconductor για την μελέτη της διαφορικής έκφρασης των δεδομένων RNA-SEQ. Η κατανομή Poisson χρησιμοποιείται για να λάβει υπόψη τόσο τη βιολογική όσο και την τεχνική μεταβλητότητα. Οι εμπειρικές μέθοδοι Bayes χρησιμοποιούνται για τον μετριασμό του βαθμού διασποράς των μεταγράφων, βελτιώνοντας την αξιοπιστία των συμπερασμάτων (Robinson et al., 2009).

2.3.11 DAVID

Η βάση δεδομένων DAVID (Database for Annotation, Visualization and Integrated Discovery) παρέχει μία σειρά από εργαλεία λειτουργικής ανάλυσης, τα οποία μπορούν να διευκολύνουν τους ερευνητές να κατανοήσουν τη βιολογική σημασία που εμπεριέχεται σε μια μακριά λίστα με γονίδια. Όλα τα εργαλεία στοχεύουν να παρέχουν λειτουργική ερμηνεία μεγάλων λιστών γονιδίων που προέρχονται από γενωμικές μελέτες, όπως για παράδειγμα μικροσυστοιχίες και πρωτεωμική.

Το DAVID αποτελείται από τη βάση δεδομένων DAVID (DAVID Knowledgebase) και πέντε ενσωματωμένες, διαδικτυακές σουίτες εργαλείων σχολιασμού: το εργαλείο ταξινόμησης (DAVID Gene Functional Classification Tool), το εργαλείο σχολιασμού DAVID (Functional Annotation Tool), το εργαλείο μετατροπής των κωδικών των γονιδίων (DAVID Gene ID Conversion Tool), επίδειξη ονομάτων των γονιδίων (DAVID Gene Name Viewer) και το φυλλομετρητή γονιδιώματος παθογόνου (DAVID NIAID Pathogen Genome Browser). Για κάθε λίστα γονιδίων, το DAVID εκτελεί εμπλουτισμό γονιδιακών όρων (gene-term enrichment analysis). Η απλή Ανάλυση Εμπλουτισμού (Single Enrichment Analysis, SEA), που εκτελεί το DAVID, περιλαμβάνει, σε δύο στάδια, τον υπολογισμό των σχετικών εμπλουτισμών και το στατιστικό έλεγχο των τιμών τους μέσω της υπεργεωμετρικής κατανομής. Για μία δεδομένη λίστα γονιδίων, τα εργαλεία του DAVID μπορούν να χρησιμοποιηθούν για τα κάτωθι: εύρεση εμπλουτισμένων γονιδιακών οντολογιών, οπτικοποίηση γονιδίων στη BioCarta & χάρτες KEGG μονοπατιών, παράθεση των συσχετιζόμενων γονιδίων, αναζήτηση λειτουργικά συσχετιζόμενων γονιδίων στη λίστα με τα DEGs, λίστα με τις αλληλεπιδρούσες πρωτεΐνες, επισήμανση των λειτουργικών domains και των μοτίβων των πρωτεϊνών, ανακατεύθυνση σε σχετική βιβλιογραφία, καθώς και άλλες λειτουργίες (Jiao et al., 2012).

2.3.12 G: PROFILER

Το g:Profiler είναι ένα δημόσιο λογισμικό για τον χαρακτηρισμό και τη λειτουργική ανάλυση λιστών γονιδίων. Το g: Profiler έχει μια απλή φιλική προς το χρήστη διεπαφή ιστού και είναι προς το παρόν διαθέσιμη για 400+ είδη, συμπεριλαμβανομένων θηλαστικών, φυτών, μυκήτων, εντόμων από την Ensembl και την Ensembl Genomes. Το g:Profiler ενημερώνεται περίπου κάθε τρεις μήνες. Αποτελείται από τα ακόλουθα εργαλεία: g:GOSt, g:Convert, g:Orth, g:SNPense.

Το g:GOSt είναι το βασικό μέρος του g: Profiler, πραγματοποιεί στατιστική ανάλυση εμπλουτισμού, με σκοπό τη λειτουργική ανάλυση των λιστών γονιδίων που παρέχονται από τους χρήστες. Η ανάλυση περιλαμβάνει στοιχεία από γονιδιακές οντολογίες, βιολογικά μονοπάτια, ρυθμιστικά μοτίβα μεταγραφικών παραγόντων και microRNA, και αλληλεπιδράσεις πρωτεϊνών.

2.3.13 KEGG

Η KEGG (Kyoto Encyclopedia of Genes and Genomes) είναι μια βάση δεδομένων, που αποσκοπεί στη συστηματική ανάλυση των λειτουργιών των γονιδίων, και συνδέει τις γονιδιωματικές πληροφορίες με λειτουργικές πληροφορίες. Περιέχει γραφικές αναπαραστάσεις βιολογικών μονοπατιών και κυτταρικών διεργασιών, όπως μεταβολισμός, μεταγωγή σήματος και κυτταρικός κύκλος. Η KEGG (Kyoto Encyclopedia of Genes and Genomes, <http://www.genome.jp/kegg/pathway.html>) αποτελεί την παλαιότερη βάση δεδομένων βιολογικών μονοπατιών. Είναι ενσωματωμένη στα περισσότερα διαδικτυακά εργαλεία λειτουργικής ανάλυσης και γι' αυτό το λόγο είναι η πιο δημοφιλής βάση βιολογικών μονοπατιών (Kanehisa & Goto, 2000).

2.3.14 CIBERSORTx

Το CIBERSORTx είναι ένα αναλυτικό εργαλείο, το οποίο δημιουργήθηκε από το εργαστήριο των Alizadeh και Newman, με σκοπό την εκτίμηση της

αφθονίας των διαφόρων κυτταρικών τύπων σε ένα μικτό κυτταρικό πληθυσμό, με τη χρήση δεδομένων γονιδιακής έκφρασης.

Το CIBERSORTx επιτρέπει στους χρήστες να επεξεργαστούν δεδομένα γονιδιακής έκφρασης, τα οποία προέρχονται από ένα μείγμα διαφορετικών κυτταρικών τύπων. Για τον καθορισμό των κυτταρικών τύπων που υπάρχουν στο δείγμα, το πρόγραμμα χρησιμοποιεί τα δεδομένα έκφρασης των γονιδίων που υπάρχουν στο δείγμα, καθώς και ένα αρχείο με γονιδιακές υπογραφές (signature matrix file), που απαριθμεί τα γονίδια, που προσδιορίζουν το προφίλ έκφρασης για κάθε κυτταρικό τύπο ενδιαφέροντος. Οι χρήστες μπορούν είτε να χρησιμοποιήσουν διαθέσιμο αρχείο γονιδιακών υπογραφών είτε να δημιουργήσουν το δικό τους. Το αρχείο με τις γονιδιακές υπογραφές του CIBERSORT, που ονομάζεται LM22, περιέχει 547 γονίδια και διακρίνει 22 διαφορετικούς ανθρώπινους φαινότυπους κυττάρων του αιμοποιητικού συστήματος, συμπεριλαμβανομένων 7 τύπων T κυττάρων 7, B κύτταρα, πλασματοκύτταρα, κύτταρα φυσικοί-φονείς, κύτταρα της μυελικής σειράς. Επίσης, δεδομένης της αυξανόμενης χρήσης του single cell RNA-SEQ, το CIBERSORTx προσφέρει επίσης τη δυνατότητα στο χρήστη να δημιουργήσει το αρχείο με τις γονιδιακές υπογραφές από δεδομένα single-cell RNA-SEQ. Για κάθε δείγμα, υπολογίζονται από τον αλγόριθμο οι αναλογίες των 22 τύπων κυττάρων του ανοσοποιητικού, μαζί με τις αντίστοιχες μετρικές, συμπεριλαμβανομένων του συντελεστή συσχέτισης Pearson και της τιμής p . Η τιμή p αντιπροσωπεύει τη στατιστική σημασία των αποτελεσμάτων σε όλους τους κυτταρικούς υποπληθυσμούς και μπορεί να χρησιμοποιηθεί για να αποκλειστούν τα αποτελέσματα με μικρότερη ακρίβεια (Newman et al., 2019).

2.3.15 X-CELL

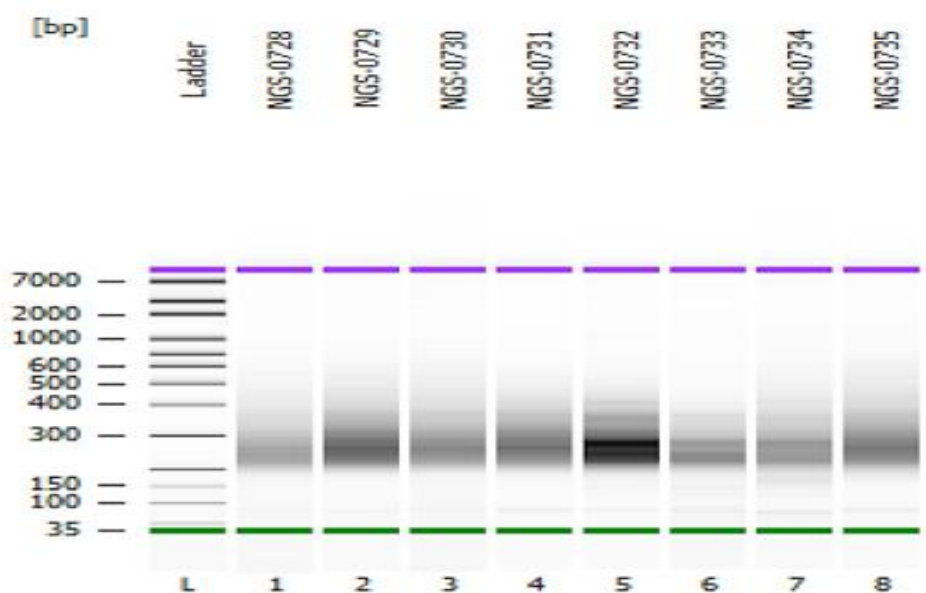
Το xCell είναι ένα διαδικτυακό εργαλείο, το οποίο εκτελεί ανάλυση εμπλουτισμού κυτταρικών τύπων. Χρησιμοποιεί δεδομένα γονιδιακής έκφρασης από 64 είδη μεσεγγυματικών και κυττάρων του ανοσοποιητικού. Το xCell είναι μία μέθοδος που βασίζεται σε γονιδιακές υπογραφές, από χιλιάδες κυτταρικούς τύπους από διάφορες πηγές. Το xCell εφαρμόζει μία νέα τεχνική

για τη μείωση των συσχετίσεων μεταξύ παρόμοιων κυτταρικών τύπων. Οι γονιδιακές υπογραφές του xCell επικυρώθηκαν με τη χρήση in-silico προσομοιώσεων και κυτταρομετρίας ανοσοφαινότυπου, και βρέθηκε να είναι ανώτερες προηγούμενων μεθόδων. Το xCell επιτρέπει στους ερευνητές να απεικονίσουν την κυτταρική ετερογένεια των προφίλ έκφρασης των ιστών (Aran et al., 2017).

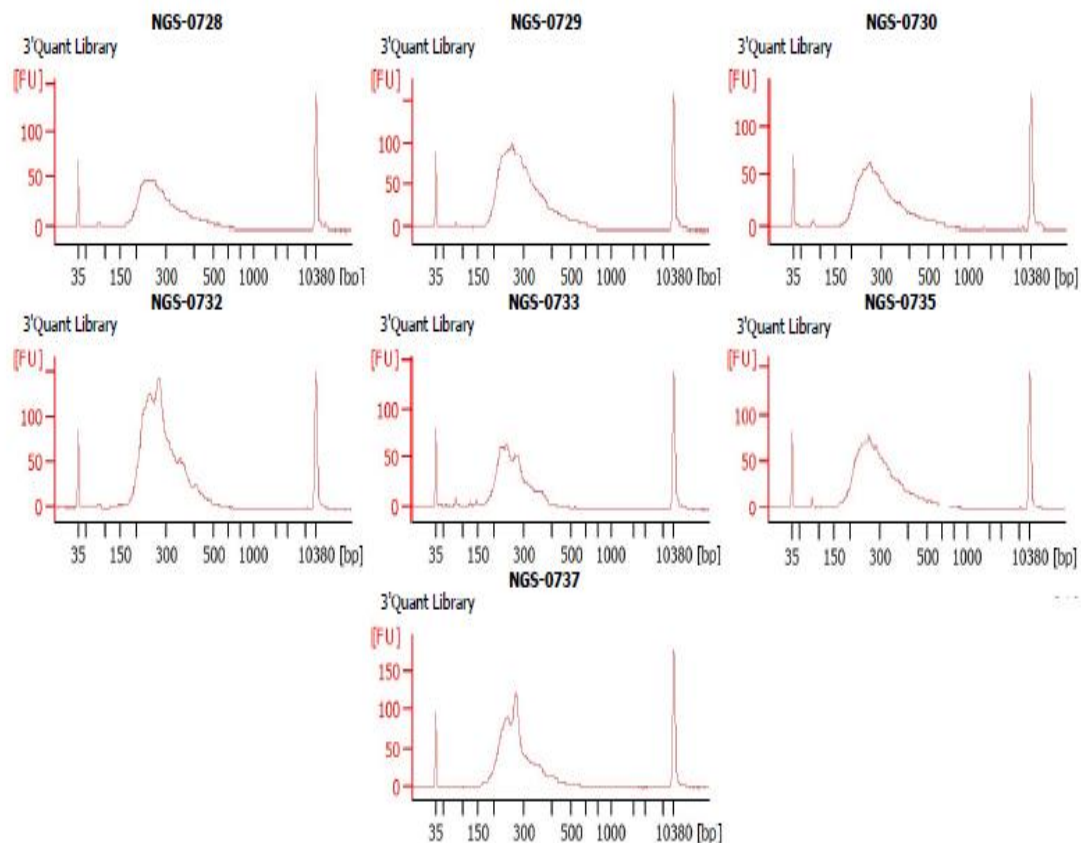
ΚΕΦΑΛΑΙΟ 3 – ΑΠΟΤΕΛΕΣΜΑΤΑ

3.1 ΕΛΕΓΧΟΣ ΠΟΙΟΤΗΤΑΣ RNA

Από τον έλεγχο της ποιότητας του RNA, που πραγματοποιήθηκε με το Bioanalyzer, κατόπιν της δημιουργίας της βιβλιοθήκης, πριν την αλληλούχιση, συμπεραίνουμε ότι η ποιότητα και ποσότητα του RNA των δειγμάτων μας είναι ικανοποιητική. Επίσης, η κατανομή του RNA στα δείγματα είναι ομοιόμορφη. Τα παραπάνω φαίνονται και στις κάτωθι εικόνες (*Εικόνα 5*, *Εικόνα 6*).



Εικόνα 5. Έλεγχος ποιότητας RNA - Ηλεκτροφόρηση



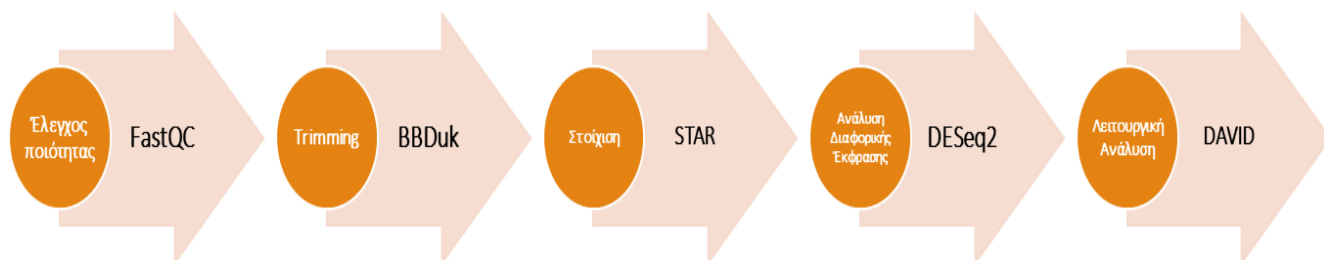
Εικόνα 6. Έλεγχος ποιότητας RNA – Διαγράμματα

3.2. ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ΡΟΗ QUANTSEQ

Η βιοπληροφορική ροή (pipeline), που αναπτύχθηκε για την ανάλυση των δεδομένων QuantSeq 3' mRNA-SEQ Library Prep Kit FWD για Illumina (Lexogen), αναγνώσεις μονής ανάγνωσης. Αποτελείται από δύο μέρη. Το πρώτο μέρος, χρησιμοποιεί το περιβάλλον Unix/Bash, και το δεύτερο μέρος τη γλώσσα προγραμματισμού R. Το πρώτο μέρος συνίσταται στον έλεγχο ποιότητας με το εργαλείο FASTQC, στο trimming, με το εργαλείο BBduk, και στη στοίχιση και τη μέτρηση των μεταγράφων (gene counts) με τον αλγόριθμο STAR, και τον αλγόριθμο Salmon. Στο δεύτερο μέρος έχει χρησιμοποιηθεί η γλώσσα προγραμματισμού R, και ειδικότερα το πακέτο DESeq2 και το πακέτο edgeR. Το δεύτερο μέρος αποτελείται από προεπεξεργασία, κανονικοποίηση, οπτικοποίηση δεδομένων και ανάλυση διαφορικής έκφρασης γονιδίων. Κατόπιν, έγινε έλεγχος εμπλουτισμού γονιδιακών οντολογιών με τη χρήση

των διαδικτυακών προγραμμάτων DAVID και g:Profiler και ψηφιακή κυτταρομετρία ροής με τα διαδικτυακά εργαλεία CIBERSORT-X και xCell.

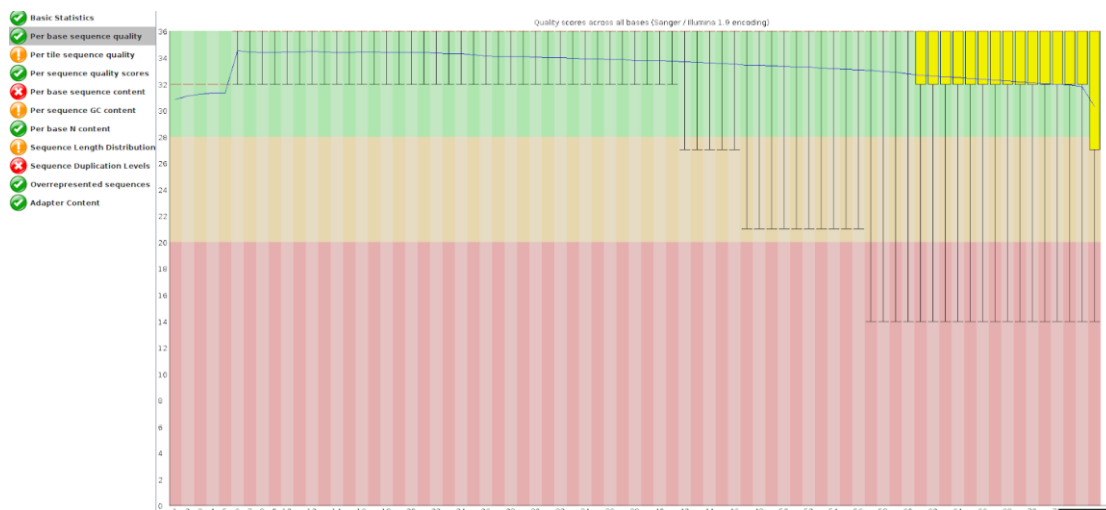
Η βιοπληροφορική ροή που κρίθηκε ποιοτικά πιο ικανοποιητική ήταν η κάτωθι: FastQC ⇒ BBDuk ⇒ STAR ⇒ DESeq2 ⇒ DAVID (Εικόνα 7):



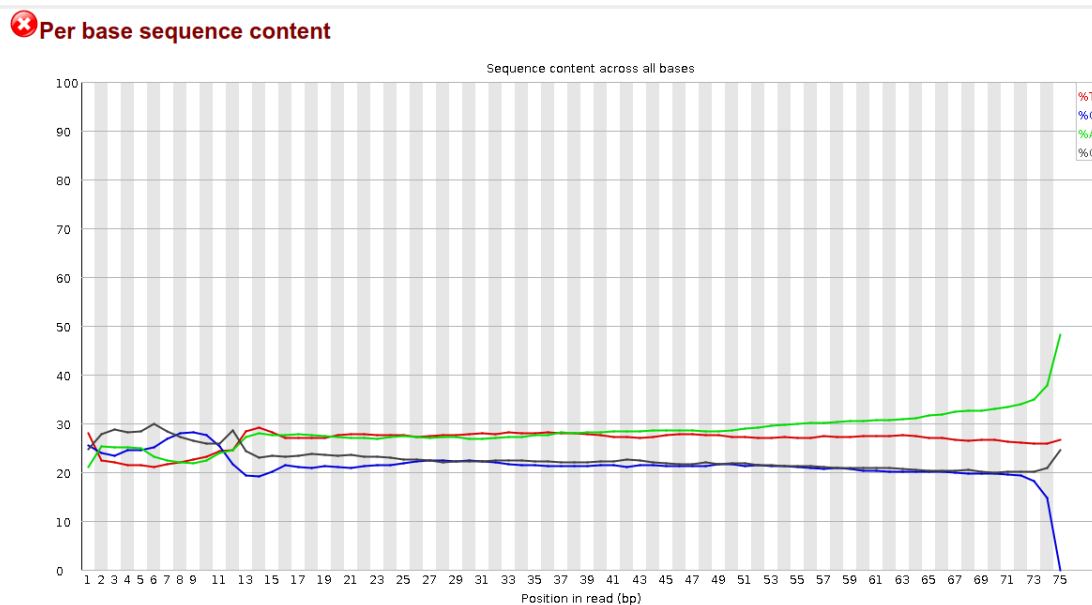
Εικόνα 7. Η βιοπληροφορική ροή QuantSeq που αναπτύχθηκε

3.2.1 ΕΛΕΓΧΟΣ ΠΟΙΟΤΗΤΑΣ

Ο έλεγχος ποιότητας πραγματοποιήθηκε με το εργαλείο FastQC. Από τις αναλύσεις που πραγματοποιήθηκαν με το εργαλείο FastQC, σε γενικές γραμμές, τα δείγματα έχουν παρόμοια ποιότητα. Οι τελευταίες βάσεις της αλληλουχίας έχουν χαμηλότερη ποιότητα (per base sequence quality < 30) (Εικόνα 8). Από το γράφημα της ανά βάση περιεχόμενο αλληλουχίας (per base sequence content), παρατηρείται ξεκάθαρη μη-μονόμορφη κατανομή των βάσεων για τα πρώτα 10-15 νουκλεοτίδια (Εικόνα 9).



Εικόνα 8. Έλεγχος ποιότητας με το εργαλείο FastQC στο δείγμα 0728 (per base sequence quality)



Εικόνα 9. Έλεγχος ποιότητας με το εργαλείο FastQC για το δείγμα 0728 (per base sequence content)

3.2.2 TRIMMING

Το trimming κρίθηκε απαραίτητο για την απομάκρυνση των προσαρμογέων (adapters), της πολύ-(A) ουράς και της χαμηλής ποιότητας άκρων της αλληλουχίας. Επιπρόσθετα, μιας και η προσέγγιση τυχαίας εκκίνησης (random priming) επίσης εισαγάγει bias στις πρώτες 12 βάσεις της αλληλουχίας, το trimming είναι αναγκαίο και για αυτό το λόγο. Το trimming πραγματοποιήθηκε με το εργαλείο BBDuk.

3.3.3 ΣΤΟΙΧΙΣΗ

Μετά το trimming, πραγματοποιήθηκε στοίχιση και ποσοτικοποίηση των μεταγράφων (gene counts) με τα εργαλεία STAR και Salmon.

- **STAR**

Σαν δείκτης του γονιδιώματος (genome index), χρησιμοποιήθηκε έτοιμος δείκτης που ανακτήθηκε από το διαδίκτυο (GRCh38_Ensembl99_

sparseD3_sjdbOverhang99)(http://labshare.cshl.edu/shares/gingeraslab/www_data/dobin/STAR/STARgenomes/Human/GRCh38_Ensembl99_sparseD3_sjdbOverhang99/). Για κάθε δείγμα, ο αλγόριθμος του STAR δημιούργησε ένα αρχείο, με το όνομα ReadsPerGene.out.tab, με τον αριθμό των μεταγράφων του δείγματος που αντιστοιχούσαν σε κάθε γονίδιο (από το γονιδίωμα αναφοράς) (πίνακας με τα gene counts) (Εικόνα 10, Εικόνα 11).

FASTQ

```

1 @NS500706:58:H25Y7BGXG:1:11101:8600:1050 1:N:0:CCGCAA
2 TTCTCNGGGCGTTTCTTTGCCGATTTTGAATGTGATTTTAAAGAGTGAAAATGAGACTATCGTTTTTATAA
3 +
4 AAAAA#EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEAEE
5 @NS500706:58:H25Y7BGXG:1:11101:7695:1052 1:N:0:CCGCAA
6 TTGGANGTTCGAAGACGATCAGATACCGCTGATTCGCCACATAAACGATGCCGACCGCGATCGCGGGCGTT
7 +
8 AAAAA#EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEAEE
9 @NS500706:58:H25Y7BGXG:1:11101:10314:1053 1:N:0:CCGCAA
10 ATTTGNAGTGACAAGACAAAGAGAAGGAGTAGAAAGGATGATTCCTTTGGCCACATTTGGTACAGTCACAT
11 +
12 /A/AA#E6E6EA//EE66EE66EE6/AA/E/AE/E//E//EEEEEEEA//EEAE/</A6/EEE/6E6
13 @NS500706:58:H25Y7BGXG:1:11101:13252:1056 1:N:0:CCGCAA
14 TGTCGCTGTCTACCCCTCACAACTGCTTACAGAAATGGATCATGTCCCTTATGTTGAGGTGACCACTTAAT
15 +
16 AAAAA#EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEAEE
17 @NS500706:58:H25Y7BGXG:1:11101:19620:1057 1:N:0:CCGCAA
18 TCTGANTGCTGCTACGGGCTTCCCTGGCACTGGGATGGAAGTACCCACTACTCTGTCTTTCTATAAGTA
19 +
20 /AA/#EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEAEE
21 @NS500706:58:H25Y7BGXG:1:11101:25986:1058 1:N:0:CCGCAA
22 TAACTTNGTCTATAGGCAAGTTGTTTTCCATATCATAATTTGATGATCTCTGTGATTCATCAATGCTAGTT
23 +
24 AAAAA#EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEAEE

```



ReadsPerGene.out.tab

1 N_unmapped	384027	384027	384027	
2 N_multimapping	1300927		1300927	
3 N_noFeature	1986357		2033102	
4 N_ambiguous	634359	251761	11864	1300927
5 ENSG00000223972	0	0	0	7695105
6 ENSG00000227232	0	0	0	
7 ENSG00000278207	0	0	0	
8 ENSG00000243485	0	0	0	
9 ENSG00000284332	0	0	0	
10 ENSG00000237613	0	0	0	
11 ENSG00000208020	0	0	0	
12 ENSG00000240361	0	0	0	
13 ENSG00000186092	0	0	0	
14 ENSG00000238009	0	0	0	
15 ENSG00000239945	0	0	0	
16 ENSG00000233750	0	0	0	
17 ENSG00000208903	0	0	0	
18 ENSG00000269981	0	0	0	
19 ENSG00000239986	0	0	0	
20 ENSG00000241860	0	0	0	
21 ENSG00000222623	0	0	0	
22 ENSG00000241599	0	0	0	
23 ENSG00000279928	0	0	0	
24 ENSG00000279457	0	0	0	
25 ENSG00000273874	0	0	0	
26 ENSG00000228463	0	0	0	
27 ENSG00000286448	0	0	0	
28 ENSG00000236079	0	0	0	
29 ENSG00000236601	0	0	0	
30 ENSG00000237094	3	3	0	
31 ENSG00000209732	0	0	0	
32 ENSG00000284733	0	0	0	
33 ENSG00000233653	0	0	0	
34 ENSG00000250575	0	0	0	
35 ENSG00000278757	0	0	0	
36 ENSG00000230021	0	0	0	
37 ENSG00000235146	0	0	0	
38 ENSG00000225972	1440	1439	1	
39 ENSG00000225630	378	378	0	
40 ENSG00000237973	20	22	2	
41 ENSG00000278791	0	0	4	
42 ENSG00000229344	62	62	0	
43 ENSG00000240409	25	25	0	
44 ENSG00000248527	1636	1629	7	
45 ENSG00000198744	17	17	0	
46 ENSG00000208663	0	0	0	
47 ENSG00000284662	0	0	0	
48 ENSG00000229376	0	0	0	
49 ENSG00000228327	0	0	0	
50 ENSG00000223181	0	0	0	

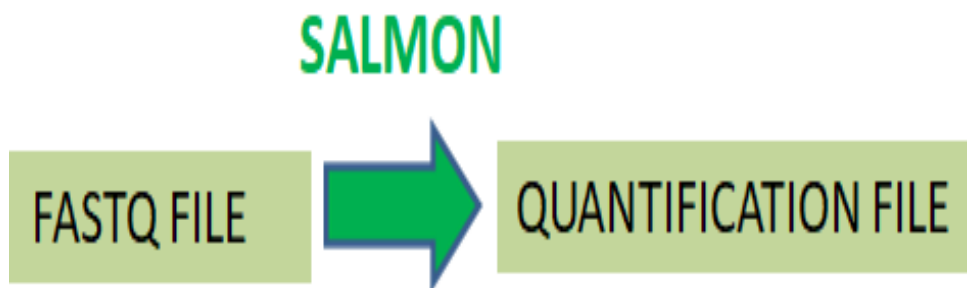
Εικόνα 10. Στοιχισμός με τον αλγόριθμο STAR

1	N_unmapped	384027	384027	384027	
2	N_multimapping	1300927		1300927	
3	N_noFeature	1986357		2033102	1300927
4	N_ambiguous	634359	251761	11864	7695105
5	ENSG000000223972		0	0	0
6	ENSG000000227232		0	0	0
7	ENSG000000278267		0	0	0
8	ENSG000000243485		0	0	0
9	ENSG000000284332		0	0	0
10	ENSG000000237613		0	0	0
11	ENSG000000268020		0	0	0
12	ENSG000000240361		0	0	0
13	ENSG000000186092		0	0	0
14	ENSG000000238009		0	0	0
15	ENSG000000239945		0	0	0
16	ENSG000000233750		0	0	0
17	ENSG000000268903		0	0	0
18	ENSG000000269981		0	0	0
19	ENSG000000239906		0	0	0
20	ENSG000000241860		0	0	0
21	ENSG000000222623		0	0	0
22	ENSG000000241599		0	0	0
23	ENSG000000279928		0	0	0
24	ENSG000000279457		0	0	0
25	ENSG000000273874		0	0	0
26	ENSG000000228463		0	0	0
27	ENSG000000286448		0	0	0
28	ENSG000000236679		0	0	0
29	ENSG000000236601		0	0	0
30	ENSG000000237094		3	3	0
31	ENSG000000269732		0	0	0
32	ENSG000000284733		0	0	0
33	ENSG000000233653		0	0	0
34	ENSG000000250575		0	0	0
35	ENSG000000278757		0	0	0
36	ENSG000000230021		0	0	0
37	ENSG000000235146		0	0	0
38	ENSG000000225972		1440	1439	1
39	ENSG000000225630		378	378	0
40	ENSG000000237973		20	22	2
41	ENSG000000278791		0	0	4
42	ENSG000000229344		62	62	0
43	ENSG000000240409		25	25	0
44	ENSG000000248527		1636	1629	7
45	ENSG000000198744		17	17	0
46	ENSG000000268663		0	0	0
47	ENSG000000284662		0	0	0
48	ENSG000000229376		0	0	0
49	ENSG000000228327		0	0	0
50	ENSG000000223181		0	0	0

Εικόνα 11. Τμήμα αρχείου ReadsPerGene.out.tab, που περιέχει τον αριθμό των μεταγράφων που αντιστοιχούν σε κάθε γονίδιο

• SALMON

Εναλλακτικά, χρησιμοποιήθηκε ο αλγόριθμος Salmon, ο οποίος βασίζεται στην ψευδοστοίχιση. Σαν δείκτης του γονιδιώματος χρησιμοποιήθηκε, δείκτης, ο οποίος κατασκευάστηκε υπολογιστικά με τη χρήση του human GRCh38 σαν το γονιδίωμα αναφοράς (Εικόνα 12).

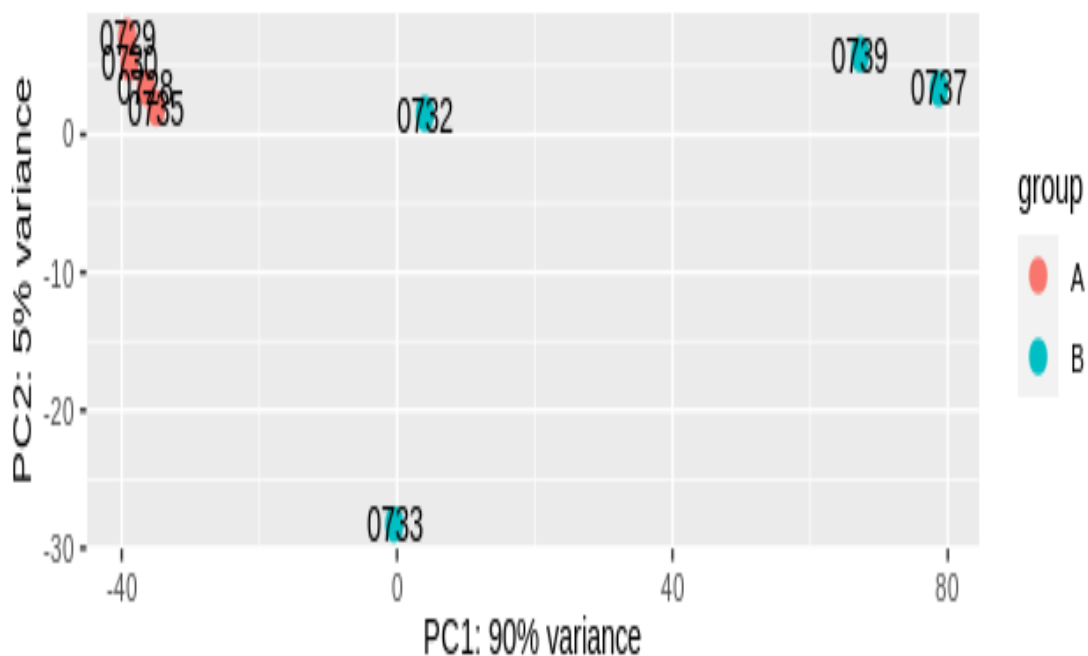


Εικόνα 12. Στοιχίση με τον αλγόριθμο Salmon

3.3.4 ΟΠΤΙΚΟΠΟΙΗΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Η οπτικοποίηση των δεδομένων πραγματοποιήθηκε με την Ανάλυση Κυρίων Συνιστωσών (Principal Component Analysis, PCA). Η PCA είναι μία από τις πιο συχνά χρησιμοποιούμενες μεθοδολογίες μείωσης της διαστασιμότητας. Η λογική πίσω από τη συγκεκριμένη μεθοδολογία βασίζεται σε απλές αρχές γραμμικής άλγεβρας. Η PCA αποσκοπεί να αναπαραστήσει τα δεδομένα σε ένα χώρο που θα αναδεικνύει τη μεγαλύτερη διασπορά τους και άρα θα συνεισφέρει στη διάκριση της ύπαρξης διακριτών ομάδων. Η PCA χρησιμοποιεί μετασχηματισμούς, που ονομάζονται κύριες συνιστώσες, και στην ουσία είναι νέες μεταβλητές, που αντιστοιχούν σε νέους άξονες και ορίζουν τη μέγιστη διασπορά των στοιχείων. Οι άξονες αντιστοιχούν στα ιδιοδιανύσματα (eigenvectors), του αρχικού συνόλου δεδομένων. Ακόμα και για δεκάδες ή εκατοντάδες μεταβλητές η συνολική διασπορά συγκεντρώνεται κατά 80% ή 90% στις πρώτες δύο ή τρεις κύριες συνιστώσες, που αρκούν για να αναπαρασταθούν με τον καλύτερο δυνατό τρόπο τα δεδομένα (Jolliffe & Cadima, 2016).

Φαίνεται από το διάγραμμα PCA που κατασκευάσαμε με τη γλώσσα προγραμματισμού R, ότι τα δεδομένα μας ομαδοποιούνται σε δύο ομάδες, που συμπίπτουν με τις ομάδες, που είχαμε διαχωρίσει αρχικά τα δεδομένα μας (Εικόνα 13).



Εικόνα 13. Διάγραμμα Ανάλυσης Κυρίων Συνιστωσών (Principal Component Analysis - PCA)

• ΑΝΑΛΥΣΗ ΔΙΑΦΟΡΙΚΗΣ ΕΚΦΡΑΣΗΣ

Η ανάλυση διαφορικής έκφρασης πραγματοποιήθηκε με τα εργαλεία DESeq2 και edgeR.

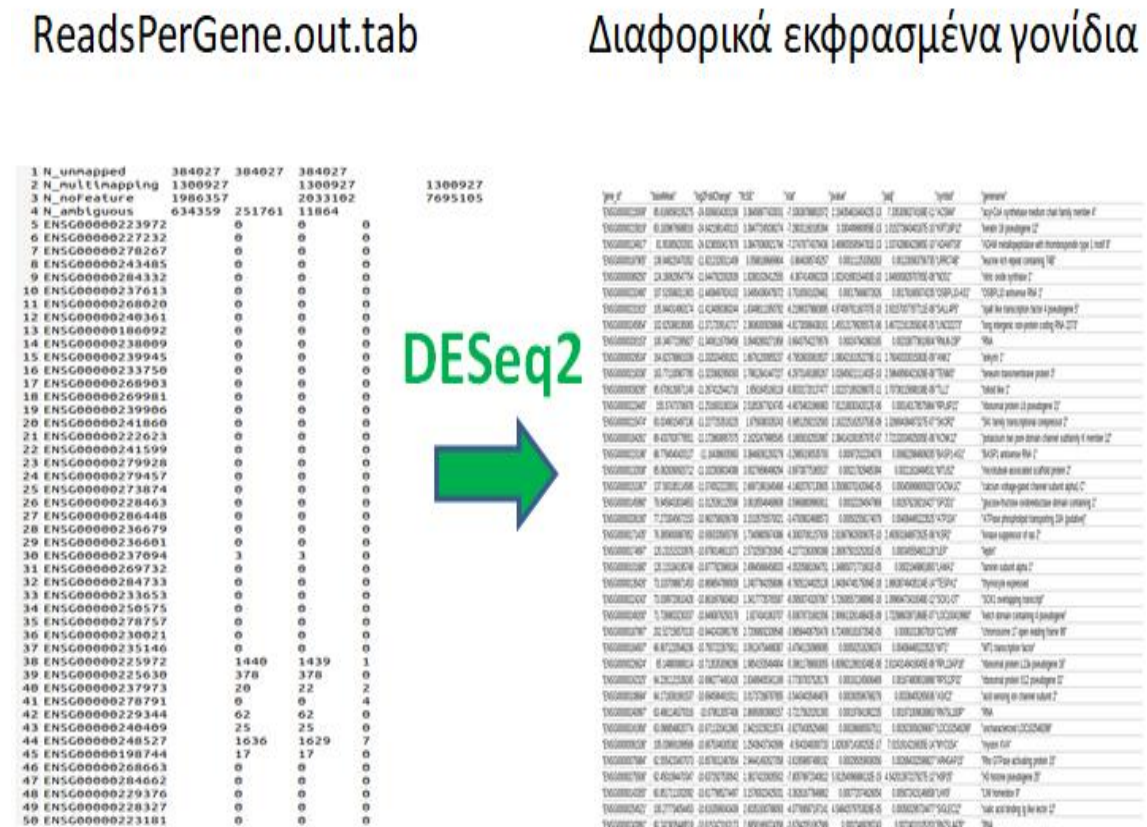
• DESEQ2

Από τη χρήση του εργαλείου DESeq2 παράχθηκε ένα αρχείο, που περιελάμβανε τα διαφορικά εκφρασμένα γονίδια μεταξύ της ομάδας A και της ομάδας B και πμές στατιστικού ελέγχου (μεταξύ άλλων η τιμή p και το \log_2 foldchange) (Εικόνα 14, Εικόνα 15).

Διπλωματική εργασία: «Ανάπτυξη βιοπληροφορικών ροών και εργαλείων για την μελέτη της ανοσοβιολογικής απόκρισης στον SARS-CoV-2, με βάση τις τεχνολογίες αλληλοτύχισης επόμενης γενιάς (NGS-RNAseq)»

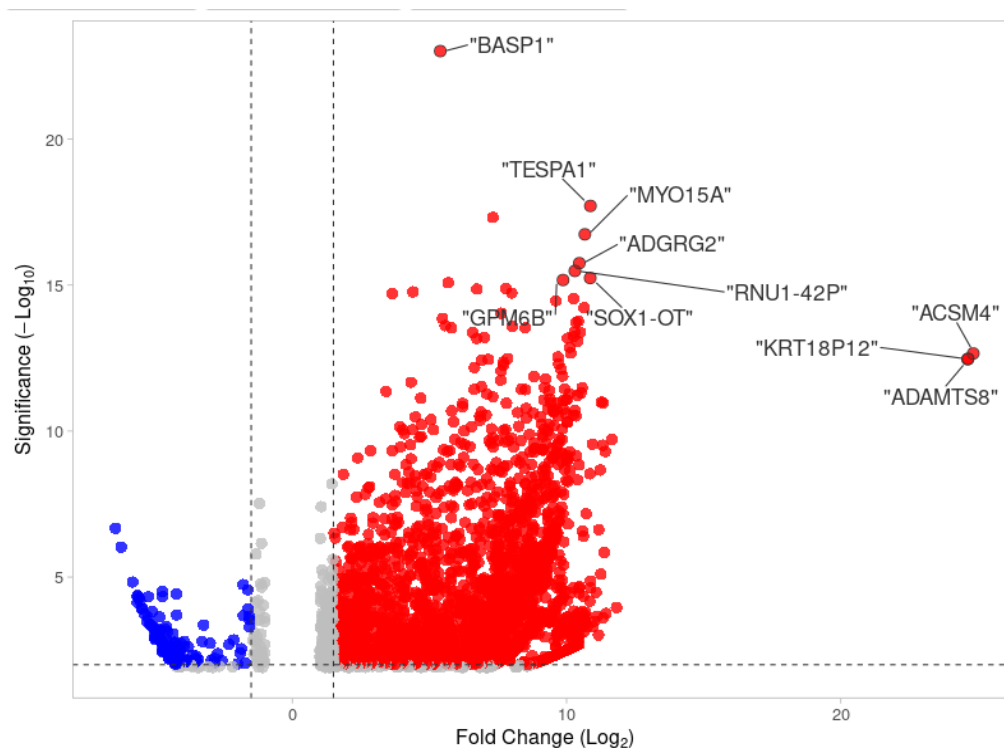
"gene_id"	"baseMean"	"log2FoldChange"	"lfcSE"	"stat"	"pvalue"	"padj"	"symbol"	"genename"
"ENSG00000215009"	95.6168590152575	-24.930634263156	3.38458977433031	-7.3963788802072	2.19435461640422E-13	7.3353090274168E-11	"ACSM4"	"acyl-CoA synthetase medium chain family member 4"
"ENSG00000215819"	83.1839676680121	24.6421981400115	3.8477245036274	-7.28031160185394	3.3304999095959E-13	1.01527384340187E-10	"KRT18P12"	"keratin 18 pseudogene 12"
"ENSG00000134917"	81.953856202001	24.6236550417878	3.38479360621794	-7.27478774379436	3.46965559654781E-13	1.03742880423985E-10	"ADAMTS8"	"ADAM metalloproteinase with thrombospondin type 1 motif 8"
"ENSG00000197905"	139.948220470352	-11.021328311409	3.0588189669904	-3.8646395745257	0.000111253358263	0.001230063756735	"LRRFC4B"	"leucine rich repeat containing 74B"
"ENSG00000208250"	124.180629547754	-11.6447922002639	1.82881026412555	-6.3674140662328	1.92241690154403E-10	1.84600082970765E-08	"NOS1"	"nitric oxide synthase 1"
"ENSG00000232490"	107.65588311903	-11.4046878242102	3.04954390475072	-3.7516931029461	0.000175668072626	0.001791665074235	"OSBP1L10-AS1"	"OSBP1L10 antisense RNA 1"
"ENSG00000201915"	105.443414963174	-11.434060360244	1.83498111950782	-6.21990379883895	4.97459781160707E-10	3.9315707787711E-08	"SALLAP5"	"spalt like transcription factor 4 pseudogene 5"
"ENSG00000249554"	100.625388195065	-11.3717209141717	2.36060009268866	-4.81730088438161	1.45513179925875E-06	3.46722161355824E-05	"LINC02273"	"long intergenic non-protein coding RNA 2273"
"ENSG00000201533"	100.346772398527	-11.34061167709458	3.09482800271958	-6.66437542279579	0.000247942883165	0.002338773610604	"RNU6-23P"	"RNA"
"ENSG00000209534"	164.823788601039	-11.3320244591821	1.66761200955237	-6.79536030819537	1.08042161052279E-11	1.76040330151083E-09	"ANKK1"	"ankyrin 1"
"ENSG00000218336"	163.771100907785	-11.3233682958393	1.79812641447227	-6.29731491885267	3.02845621111402E-10	2.5964956041628E-08	"TENM3"	"teneurin transmembrane protein 3"
"ENSG00000308295"	95.670613067149	-11.2674125441716	1.6561645166118	-8.03331720137477	1.02237159028807E-11	1.70736115698108E-09	"TLL1"	"tolloid like 1"
"ENSG00000223445"	155.5743706978	-11.169310831064	2.51853977624745	-4.46754631966983	7.91218830342012E-06	0.000140178575984	"RPL6P21"	"ribosomal protein L6 pseudogene 21"
"ENSG00000215474"	93.0249015497136	-11.2277253516225	1.8759380332543	-4.96512592152583	2.16222516293753E-09	1.32669438487327E-07	"SKOR2"	"SKI family transcriptional repressor 2"
"ENSG00000184261"	89.4337926779551	-11.1728608957075	2.1625479985545	-5.16658162553987	2.38414109195797E-07	7.72132034925005E-06	"KCNK12"	"potassium two pore domain channel subfamily K member 12"
"ENSG00000215196"	88.7794540420127	-11.1643860505993	3.8468381293279	-3.29850190535793	0.000972022240078	0.006822984806035	"BASP1-AS1"	"BASP1 antisense RNA 1"
"ENSG00000132938"	85.0828390920712	-11.1023938034088	3.00276696498294	-3.69738775365537	0.000217829485394	0.00211818494531	"MTUS2"	"microtubule associated scaffold protein 2"
"ENSG00000151067"	137.593185114595	-11.0745222228551	2.669719816145468	-4.14820761730065	3.35088370242064E-05	0.000459998003928	"CACNA1C"	"calcium voltage-gated channel subunit alpha1 C"
"ENSG000001304653"	79.948433034653	-11.0125391125596	3.06185546468609	-3.5966809960911	0.000322294547969	0.002876239216427	"GFOD1"	"glucose-fructose oxidoreductase domain containing 1"
"ENSG00000206190"	77.273049672153	-10.97358266789	3.1510765570021	-3.47939024688573	0.000502256174079	0.004084465223255	"ATP10A"	"ATPase phospholipid transporting 10A (putative)"
"ENSG00000171435"	76.389900887852	-10.9393353655795	1.73490805674386	-6.30837081157939	2.81987962930907E-10	4.49393184897262E-08	"KSR2"	"kinase suppressor of ras 2"
"ENSG00000174697"	120.23151523978	-10.8790146611073	2.5732587263645	-4.22772363090388	2.36067501525281E-05	0.000343554601128	"LEP"	"leptin"
"ENSG00000101807"	120.115184195748	-10.877823989184	2.49945666458033	-4.35205881064751	1.3486507171716E-05	0.00021949861189	"LAMA1"	"laminin subunit alpha 1"
"ENSG00000135426"	73.103708871453	-10.8698547890399	1.24077842058086	-8.76051244025128	1.94364748175094E-18	1.86638749435134E-14	"TESP1"	"thymocyte expressed"
"ENSG00000204243"	73.0399720610428	-10.861897634819	1.3417735765587	-8.0590743287067	5.72808551798895E-16	1.0996847341048E-12	"SOX1-OT"	"SOX1 overlapping transcript"
"ENSG00000249207"	71.729832320337	-10.8490876250178	1.8274341033707	-5.93678731681556	2.9068132514884E-09	1.72288803971888E-07	"LOC100419960"	"ketch domain containing 4 pseudogene"
"ENSG00000187987"	202.527156570133	-10.8442433991765	2.7206883239548	-3.98584406750478	6.72466618187354E-05	0.00081013607919	"C1orf99"	"chromosome 17 open reading frame 99"
"ENSG00000184937"	66.9071220546236	-10.755722878911	3.09124754488387	-3.47941159898095	0.000502516296374	0.004084465223255	"WT1"	"WT1 transcription factor"
"ENSG00000228624"	65.14880889114	-10.713535096286	1.98541535494904	-3.961178683055	6.80982128919248E-08	2.61043149419045E-06	"RPL13AP16"	"ribosomal protein L13a pseudogene 16"
"ENSG00000243255"	64.2851121539245	-10.6962774491426	2.8349405341169	-3.73707873581179	0.000161245606469	0.001674809016898	"RPS12P31"	"ribosomal protein S12 pseudogene 31"
"ENSG00000108694"	74.183911991537	-10.694596195131	3.0173288787855	-3.5454303544698	0.000393596788276	0.0033640265636	"ASIC2"	"acid sensing ion channel subunit 2"
"ENSG00000148379"	63.4851148379316	-10.679613057406	2.86595093690157	-3.72175620291337	0.000197841922235	0.00197100936863	"RNTS1B3P"	"RNA"
"ENSG00000241956"	63.098548027074	-10.671320412865	2.94210239213574	-3.62704305254993	0.000286685057511	0.002623056299067	"LOC102546299"	"uncharacterized LOC102546299"
"ENSG0000019536"	105.03969109569	-10.667346305382	1.25496437342699	-8.5043348300733	1.82639714180252E-17	7.0151914216635E-14	"MYO15A"	"myosin XVa"
"ENSG0000075884"	62.555423407073	-10.6578012487854	2.94441492627058	-3.61958897489192	0.000290596983856	0.00268432598827	"ARHGAP15"	"Rho GTPase activating protein 15"
"ENSG00000243255"	62.4501094470347	-10.6372927340812	1.36274220835502	-7.80579972340812	5.91254096866132E-15	4.5420139277927E-12	"H3P25"	"H3 histone pseudogene 25"
"ENSG00000143355"	60.851711002092	-10.6177995274497	3.15760002425031	-3.3626167748862	0.000772074626054	0.005672415146658	"LHX9"	" LIM homeobox 9"
"ENSG00000254521"	100.277740544531	-10.613599343439	2.60351930788093	-4.07769587197141	4.5484207953828E-05	0.00058329524477	"SIGLEC12"	"sialic acid binding Ig like lectin 12"
"ENSG00000243991"	60.7423635449519	-10.6152473163173	2.88581669374358	-3.67842051067589	0.00023468280743	0.0022401052093	"RNTS1A47P"	"RNA"

Εικόνα 14. Αρχείο με τα διαφορετικά εκφρασμένα γονίδια - Εργαλείο DESeq2



Εικόνα 15. Ανάλυση Διαφορικής Έκφρασης με το εργαλείο DESeq2

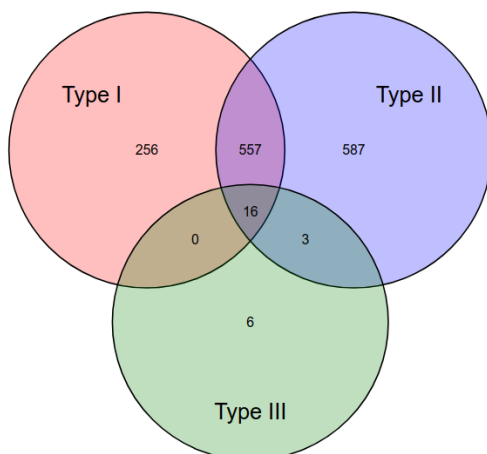
Στην ανάλυση διαφορικής έκφρασης, με το εργαλείο DESeq2 βρέθηκε ότι 4464 γονίδια ήταν διαφορετικά εκφρασμένα σε επίπεδο σημαντικότητας 5%, $|\log_2\text{foldchange}| > 1$, στην ομάδα A έναντι της ομάδας B (ελέγχου). Από αυτά τα γονίδια, 2733 γονίδια ήταν υπερεκφρασμένα στην ομάδα A σε σχέση με την ομάδα B, ενώ τα υπόλοιπα (1731) ήταν υποεκφρασμένα. Το γονίδιο με τη μεγαλύτερη διαφορά στην έκφραση μεταξύ των δύο ομάδων καθώς και με τη μεγαλύτερη διορθωμένη τιμή p ήταν το acyl-CoA synthetase medium chain family member 4 ($\log_2\text{foldchange} = 24.83066343$, διορθωμένη τιμή $p = 1.86180300485987E-19$). Κατόπιν, ακολουθούν τα γονίδια keratin 18 pseudogene 12 ($\log_2\text{foldchange} = 24.642198$) και ADAM metalloproteinase with thrombospondin type 1 motif 8 ($\log_2\text{foldchange} = 24.62365504$) (Εικόνα 16).



Εικόνα 16. Volcano plot των διαφορετικά εκφρασμένων γονιδίων που προέκυψαν με το εργαλείο DESeq2

Τα διαφορετικά εκφρασμένα γονίδια, τα οποία ρυθμίζονται από τις ιντερφερόνες, υπολογίστηκαν, με τη βοήθεια της βάσης δεδομένων Interferome, η οποία είναι μία βάση δεδομένων, που διευκολύνει την ταυτοποίηση των γονιδίων, που ρυθμίζονται από τις ιντερφερόνες (Rusinova

et al., 2013). Βρέθηκε ότι 1425 γονίδια από τη λίστα των DEGs ήταν ρυθμιζόμενα από ιντερφερόνες (Εικόνα 17).



Εικόνα 17. Διάγραμμα Venn που απεικονίζει τον αριθμό των διαφορετικά εκφρασμένων γονιδίων που ρυθμίζονται από ένα ή περισσότερους τύπους ιντερφερόνης

- **EDGER**

Με τη χρήση του edgeR παράχθηκε αρχείο με τα διαφορικά εκφρασμένα γονίδια και μετρικές στατιστικής (πιμή p, log₂foldchange) (Εικόνα 18).

GENE	logFC	logCPM	LR	PValue	FDR
ENSG00000189190	4.03964314853384	4.90987332317201	22.4210056982289	2.18966077575136E-06	0.022152442964161
ENSG00000115009	-1.20472789744161	7.60188281997912	20.8214901636326	5.04140945596553E-06	0.022152442964161
ENSG00000042429	3.28152025550094	5.63163665830226	20.3330006776316	6.50685956349213E-06	0.022152442964161
ENSG00000198242	-0.682564700887497	10.3635631323421	19.9826994688891	7.81460198047832E-06	0.022152442964161
ENSG00000238297	2.8791324814649	5.03821529525387	18.3372894663978	1.85049423673529E-05	0.040728461205088
ENSG00000157306	3.81513801129102	4.15075215846957	18.047039491739	2.15513508449182E-05	0.040728461205088
ENSG00000232234	4.28389119264172	4.59426600855577	16.7606970450513	4.240253604067E-05	0.053396010417568
ENSG00000279066	2.99251073648489	3.86726168746708	16.7086583129482	4.35816420599335E-05	0.053396010417568
ENSG00000260093	4.61819877793376	4.15232659857892	16.6819154507876	4.42003549771899E-05	0.053396010417568
ENSG00000186197	6.10876479862487	5.33217970796999	16.5617787109196	4.70905815482568E-05	0.053396010417568
ENSG00000035664	2.92201044302694	4.7195159986105	16.0836919943661	6.06036842925501E-05	0.056033702390823
ENSG00000076356	4.03495003572351	5.08894771092647	15.8731734284108	6.77320294550431E-05	0.056033702390823
ENSG00000165434	2.65161120223995	5.40059605211728	15.7284281773501	7.31169891048511E-05	0.056033702390823
ENSG00000133121	3.55521883829372	5.72833055981681	15.698049952065	7.43007332743984E-05	0.056033702390823
ENSG00000175575	2.68763279194012	5.60661119204602	15.3165025855291	9.09187808751448E-05	0.056033702390823
ENSG00000182648	4.33761526427102	5.42153477684568	15.293908895093	9.2012729905824E-05	0.056033702390823
ENSG00000127589	3.04164877313553	5.52854395578721	15.2557213031572	9.38919080541177E-05	0.056033702390823
ENSG00000142627	-1.85324438261542	5.33050926722744	15.0343876184154	0.000105569925685	0.058316330486074
ENSG00000247809	3.05791569459758	6.20563743803994	14.991391722775	0.000108002728654	0.058316330486074
ENSG00000155849	3.80108914506145	4.72140622253827	14.7497478825951	0.000122763940591	0.062500613476269
ENSG00000250098	3.96568474789493	3.50954320697547	14.5826674204363	0.000134142716767	0.062500613476269
ENSG00000279159	3.1359828568353	5.48830480022514	14.5451792260207	0.000136838326098	0.062500613476269
ENSG00000183666	2.72414145405567	3.02263621099116	14.5319834185243	0.000137800100265	0.062500613476269
ENSG00000147889	-0.910378660175153	6.61066768592618	14.0473649883132	0.000178263430903	0.075133100976046
ENSG00000225489	2.26185508670294	4.85922084827661	13.9823625332336	0.000184533589443	0.075133100976046
ENSG00000127946	2.09277280134993	5.1960331779328	13.9722361524069	0.000185530190257	0.075133100976046
ENSG00000100823	-0.802816564954502	6.95148801053977	13.8710902209521	0.000195787083096	0.076552749490413
ENSG00000115459	2.46894649713569	4.2285325602402	13.7193333093522	0.000212258202423	0.079177111346969
ENSG00000167863	-0.833967635787633	6.9550798798976	13.6304397018786	0.00022254775299	0.079177111346969
ENSG00000078246	1.41790999977239	5.94965901728788	13.5699858620339	0.000229831110968	0.079177111346969
ENSG00000280381	3.52522533010895	3.16141838660461	13.5343068582874	0.000234241879375	0.079177111346969
ENSG00000227533	5.08901299391597	4.70441135716867	13.5090742786487	0.000237412627727	0.079177111346969
ENSG00000179743	3.06785332238416	5.02812902646411	13.4277671953789	0.000247927029042	0.080321273780331
ENSG00000262454	3.76556134259185	3.99891909652251	13.3582695617795	0.000257285595737	0.080384462292735
ENSG00000186470	3.70007267845021	4.84491937742592	13.264101726356	0.000270537059969	0.080384462292735
ENSG00000197168	3.76564396753816	5.40675352891454	13.1735863177619	0.000283922838735	0.080384462292735
ENSG00000205181	3.54461448565706	4.20203308935748	13.1200550065811	0.000292151456787	0.080384462292735

Εικόνα 18. Διαφορικά εκφρασμένα γονίδια - edgeR

3.3.6 ΛΕΙΤΟΥΡΓΙΚΗ ΑΝΑΛΥΣΗ

• ΕΛΕΓΧΟΣ ΕΜΠΛΟΥΤΙΣΜΟΥ ΓΟΝΙΔΙΑΚΩΝ ΟΝΤΟΛΟΓΙΩΝ

Ο έλεγχος εμπλουτισμού των γονιδιακών οντολογιών πραγματοποιήθηκε με το πρόγραμμα DAVID και το πρόγραμμα g: Profiler.

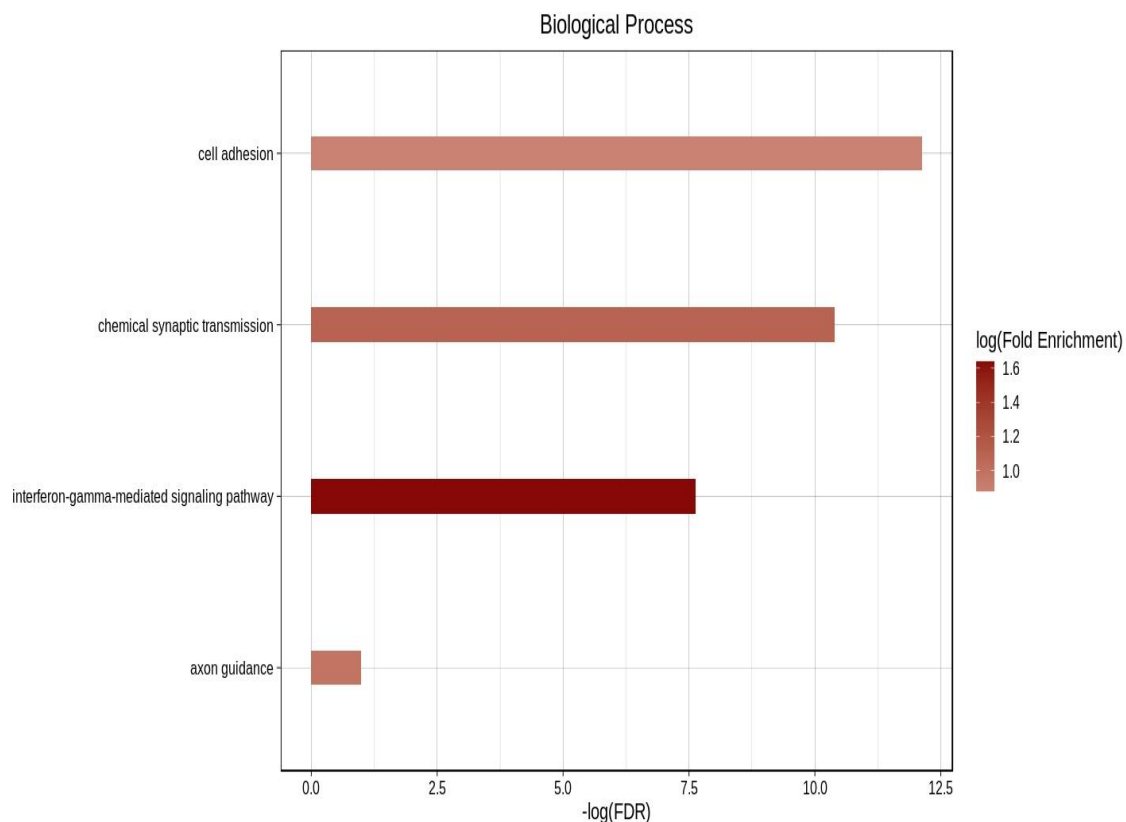
❖ DAVID

Τα αποτελέσματα από το πρόγραμμα DAVID αναφέρονται κάτωθι (*Εικόνα 19, Εικόνα 20*):

Γονιδιακές οντολογίες - Βιολογικές διαδικασίες (GO - biological process)

Οι γονιδιακές οντολογίες - βιολογικές διαδικασίες που βρέθηκαν εμπλουτισμένες σε επίπεδο σημαντικότητας 5% είναι οι εξής:

1. GO:0007155~κυτταρική προσκόλληση (cell adhesion) (διορθωμένη τιμή $p = 2,24E-04$)
2. GO:0007268~χημική συναπτική μετάδοση (chemical synaptic transmission) (διορθωμένη τιμή $p = 7,44E-04$)
3. GO:0060333~μονοπάτι σηματοδότησης διαμεσολαβούμενο από ιντερφερόνη γάμμα (interferon-gamma-mediated signaling pathway) (διορθωμένη τιμή $p = 0,005027432652$)

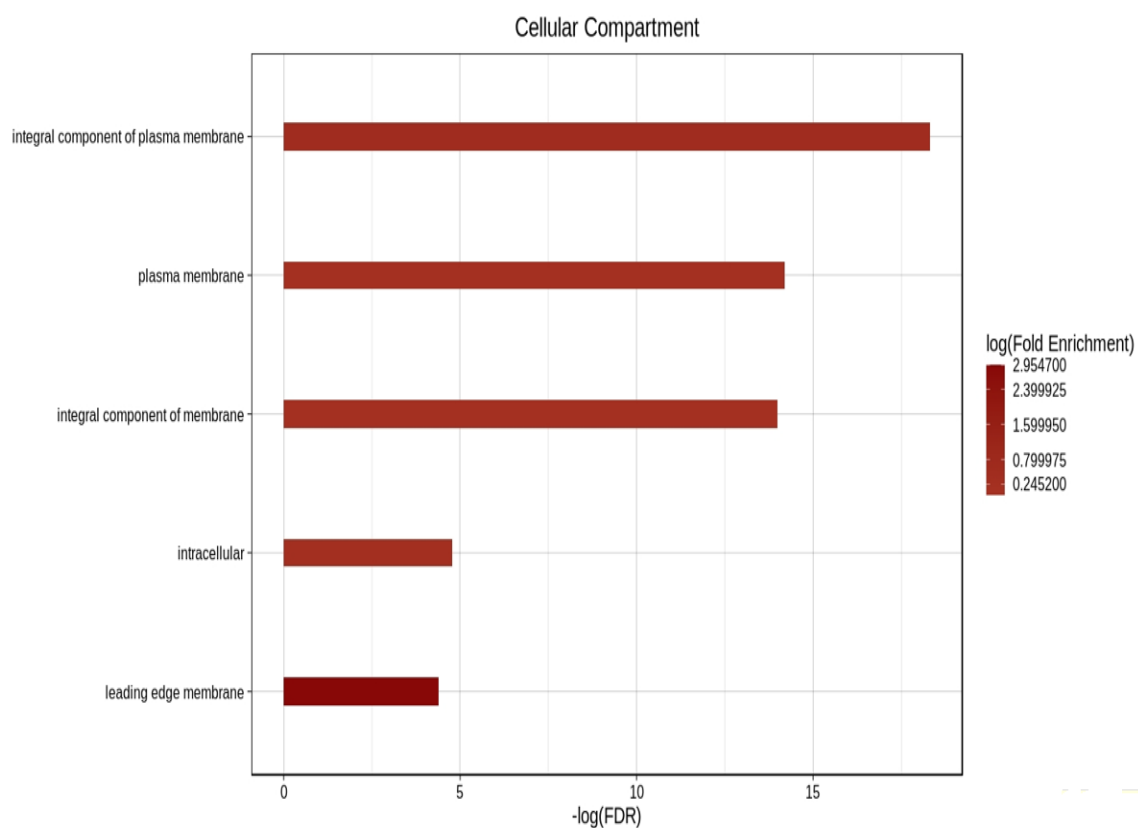


Εικόνα 19. Γονιδιακές οντολογίες – βιολογικές διαδικασίες

Γονιδιακές οντολογίες - Κυτταρικό διαμέρισμα (GO - cellular compartment)

Οι γονιδιακές οντολογίες - κυτταρικό διαμέρισμα που βρέθηκαν εμπλουτισμένες σε επίπεδο σημαντικότητας 5% είναι οι εξής:

1. GO:0005887~αναπόσπαστο συστατικό της πλασματικής μεμβράνης (integral component of plasma membrane) (διορθωμένη τιμή $p = 3,07E-0$)
2. GO:0005886~πλασματική μεμβράνη (plasma membrane) (διορθωμένη τιμή $p = 5,31E-05$)
3. GO:0016021~αναπόσπαστο κομμάτι της μεμβράνης (integral component of membrane) (διορθωμένη τιμή $p = 6,15E-05$)
4. GO:0005622~ενδοκυττάριο (intracellular) (διορθωμένη τιμή $p = 0,03647744941$)
5. GO:0031256~μεμβράνη (leading edge membrane) (διορθωμένη τιμή $p = 0,04755240441$)



Εικόνα 20. Γονιδιακές οντολογίες- κυτταρικό διαμέρισμα

Γονιδιακές οντολογίες - Μοριακή λειτουργία (GO - molecular function)

Οι γονιδιακές οντολογίες - μοριακή λειτουργία που βρέθηκαν εμπλουτισμένες σε επίπεδο σημαντικότητας 5% είναι οι εξής:

1. GO:0046872~σύνδεση σε μεταλλικά ιόντα (metal ion binding) (διορθωμένη τιμή $p = 0,02860130773$)

- **G: PROFILER**

Τα αποτελέσματα από το πρόγραμμα g: Profiler είναι παρακάτω:

Γονιδιακές οντολογίες - Βιολογικές διαδικασίες (GO - biological process)

Δεν βρέθηκαν εμπλουτισμένες γονιδιακές οντολογίες - βιολογικές διαδικασίες σε επίπεδο σημαντικότητας 5%.

Γονιδιακές οντολογίες - Κυτταρικό διαμέρισμα (GO - cellular compartment)

Οι γονιδιακές οντολογίες - κυτταρικό διαμέρισμα που βρέθηκαν εμπλουτισμένες σε επίπεδο σημαντικότητας 5% είναι οι εξής:

1. GO:0031226~intrinsic component of plasma membrane (διορθωμένη τιμή $p = 0.000004883669601$)
2. GO:0060089~integral component of plasma membrane (διορθωμένη τιμή $p = 0.00001282780293$)
3. GO:0034703~cation channel complex (διορθωμένη τιμή $p = 0.02618181815$)
4. GO:0034702~ion channel complex (διορθωμένη τιμή $p = 0.02708893543$)
5. GO:1902495~transmembrane transporter complex (διορθωμένη τιμή $p = 0.03179092277$)

Γονιδιακές οντολογίες - Μοριακή λειτουργία (GO - molecular function)

Οι γονιδιακές οντολογίες - μοριακή λειτουργία που βρέθηκαν εμπλουτισμένες σε επίπεδο σημαντικότητας 5% είναι οι εξής:

1. GO:0038023~δραστηριότητα υποδοχέα σηματοδότησης (signaling receptor activity) (διορθωμένη τιμή $p = 0.0006584738165$)
2. GO:0060089~δραστηριότητα μοριακού μετατροπέα (molecular transducer activity) (διορθωμένη τιμή $p = 0.0006584738165$)
3. GO:0004888~δραστηριότητα του διαμεμβρανικού υποδοχέα σηματοδότησης (transmembrane signaling receptor activity) (διορθωμένη τιμή $p = 0.0006782214984$)
4. GO:0004930~ δραστηριότητα υποδοχέα συζευγμένη με πρωτεΐνη G (G protein-coupled receptor activity) (διορθωμένη τιμή $p = 0.001923190985$)

5. GO:0046873~δραστηριότητα διαμεμβρανικού μεταφορέα μεταλλικών ιόντων (metal ion transmembrane transporter activity) (διορθωμένη τιμή $p = 0.03265363258$)
6. GO:0022836~δραστηριότητα καναλιών (gated channel activity) (διορθωμένη τιμή $p = 0.04036263988$)
7. GO:0005216~δραστηριότητα καναλιών ιόντων (ion channel activity) (διορθωμένη τιμή $p = 0.000985762501828$)

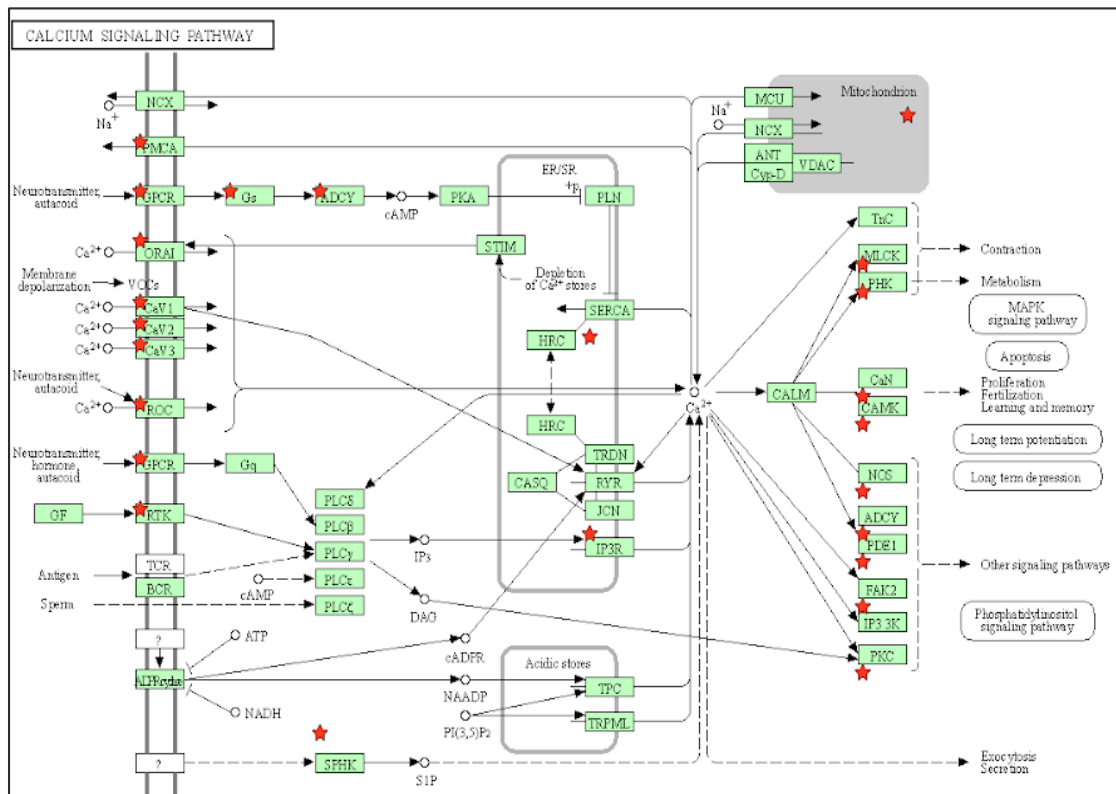
• ΕΛΕΓΧΟΣ ΕΜΠΛΟΥΤΙΣΜΟΥ ΒΙΟΛΟΓΙΚΩΝ ΜΟΝΟΠΑΤΙΩΝ

Η KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa & Goto, 2000) αποτελεί την παλαιότερη βάση δεδομένων βιολογικών μονοπατιών. Περιέχει ίσως το μεγαλύτερο όγκο πληροφορίας και είναι, από άποψη κατηγοριών, η πληρέστερη βάση. Λόγω ακριβώς της παλαιότητάς της είναι ενσωματωμένη στα περισσότερα διαδικτυακά εργαλεία λειτουργικής ανάλυσης.

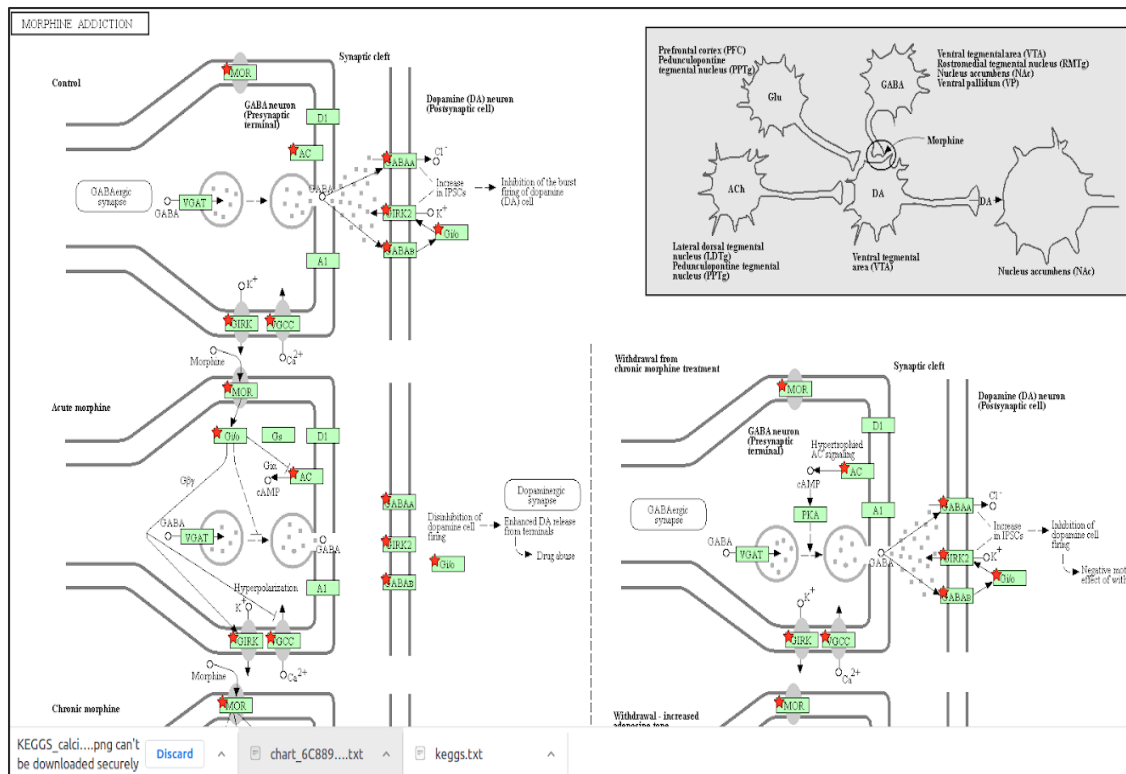
❖ DAVID

Τα Kegg's μονοπάτια που βρέθηκαν στατιστικά σημαντικά, με τη χρήση του προγράμματος DAVID (διορθωμένη τιμή $p < - 0.05$) είναι τα κάτωθι (Εικόνα 21 – Εικόνα 25):

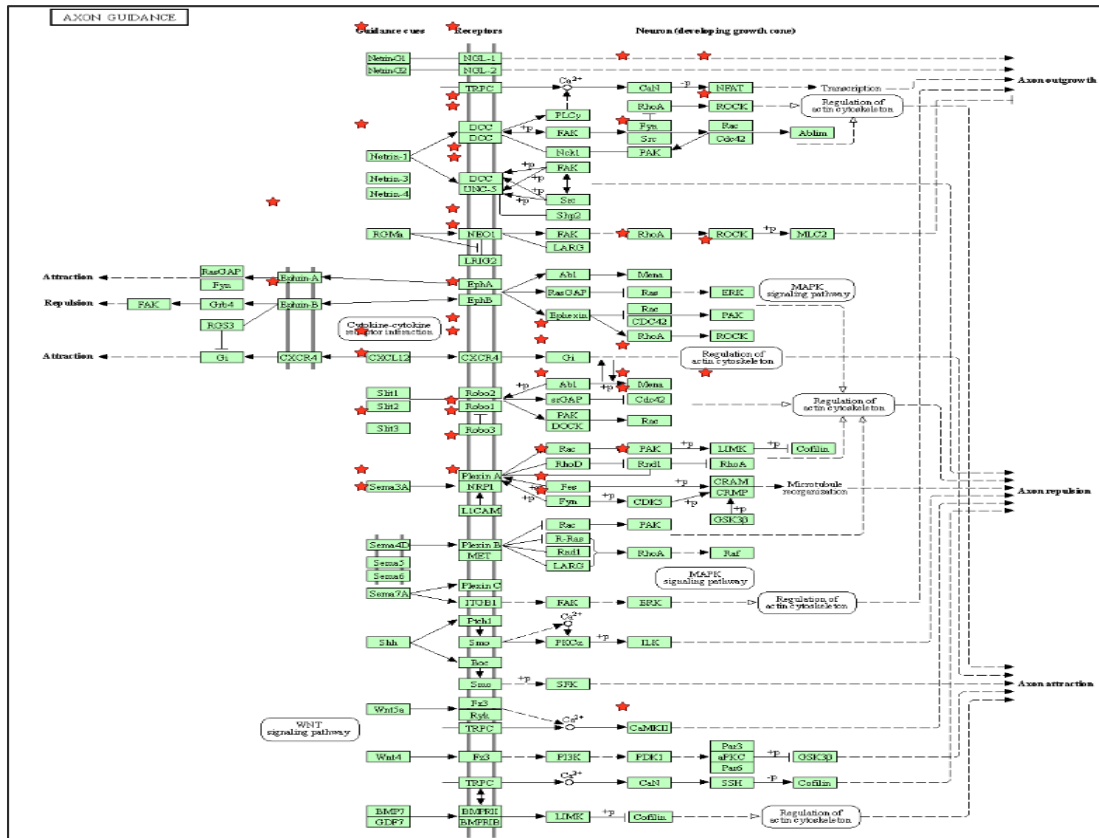
1. hsa04713: Μονοπάτι σηματοδότησης ασβεστίου (Calcium signaling pathway) (διορθωμένη τιμή $p = 0.0025388334831623455$)
2. hsa05032: Εξάρτηση από τη μορφίνη (Morphine addiction) (διορθωμένη τιμή $p = 0.005228899833199824$)
3. hsa04360: Καθοδήγηση του άξονα των νευρικών κυττάρων (Axon guidance) (διορθωμένη τιμή $p = 0.005228899833199824$)
4. hsa04514: Μόρια κυτταρικής προσκόλλησης (Cell adhesion molecules, CAMs)(διορθωμένη τιμή $p = 0.02609279464513021$)
5. hsa04380: Διαφοροποίηση των οστεοκλαστών (Osteoclast differentiation) (διορθωμένη τιμή $p = 0.03413200781928241$)



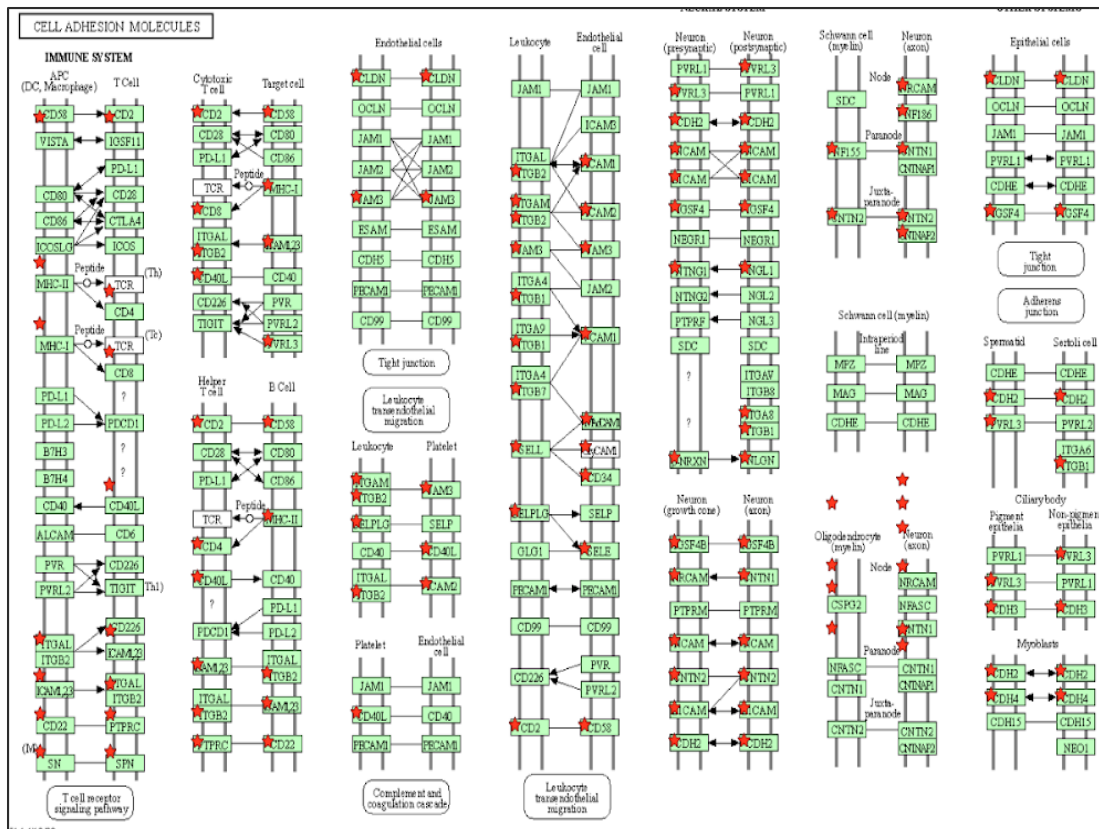
Εικόνα 21. Μονοπάτι σηματοδότησης ασβεστίου (Calcium signaling pathway)



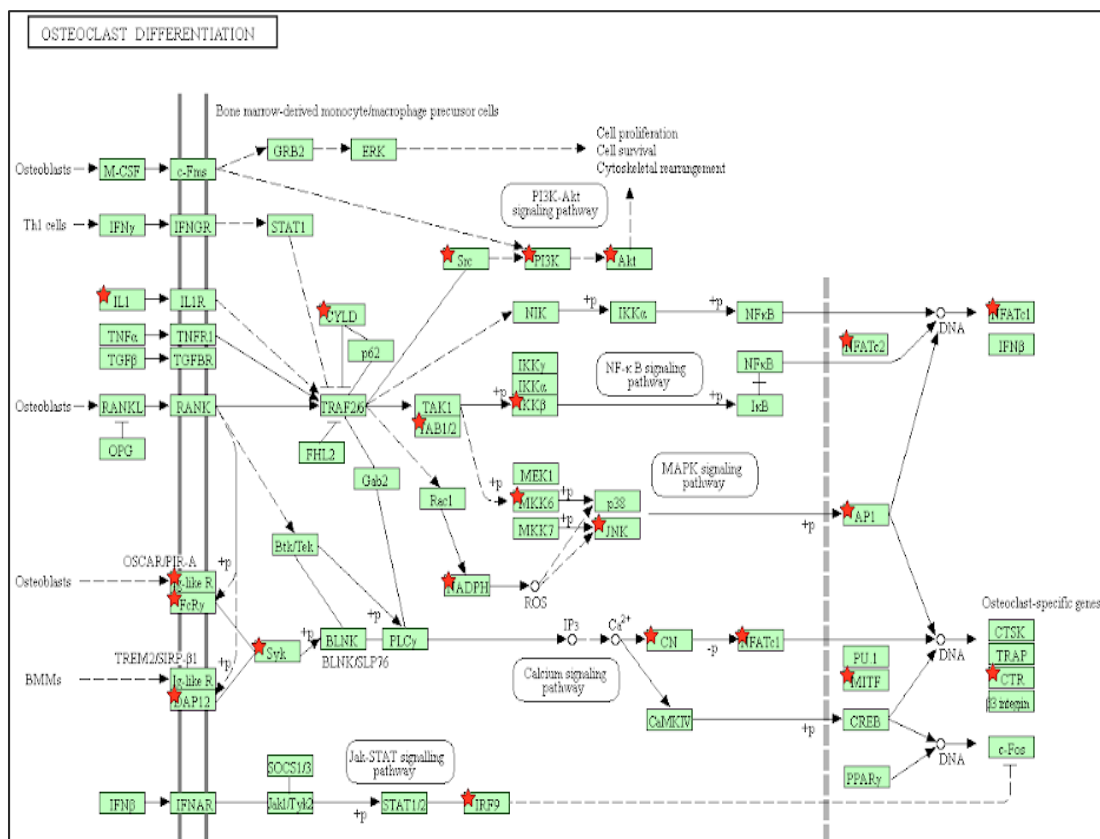
Εικόνα 22. Εξάρτηση από τη μορφίνη (Morphine addiction)



Εικόνα 23. Καθοδήγηση του άξονα των νευρικών κυττάρων (Axon guidance)



Εικόνα 24. Μόρια κυτταρικής σύντηξης (Cell adhesion molecules, CAMs)



Εικόνα 25. Διαφοροποίηση των οστεοκλαστών (Osteoclast differentiation)

❖ G:PROFILER

Τα KEGGs μονοπάτια που βρέθηκαν στατιστικά σημαντικά με τη χρήση του προγράμματος g:Profiler είναι τα κάτωθι:

1. KEGG:04020~ Μονοπάτι σηματοδότησης ασβεστίου (Calcium signaling pathway) (διορθωμένη τιμή $p = 0.02172228654$)

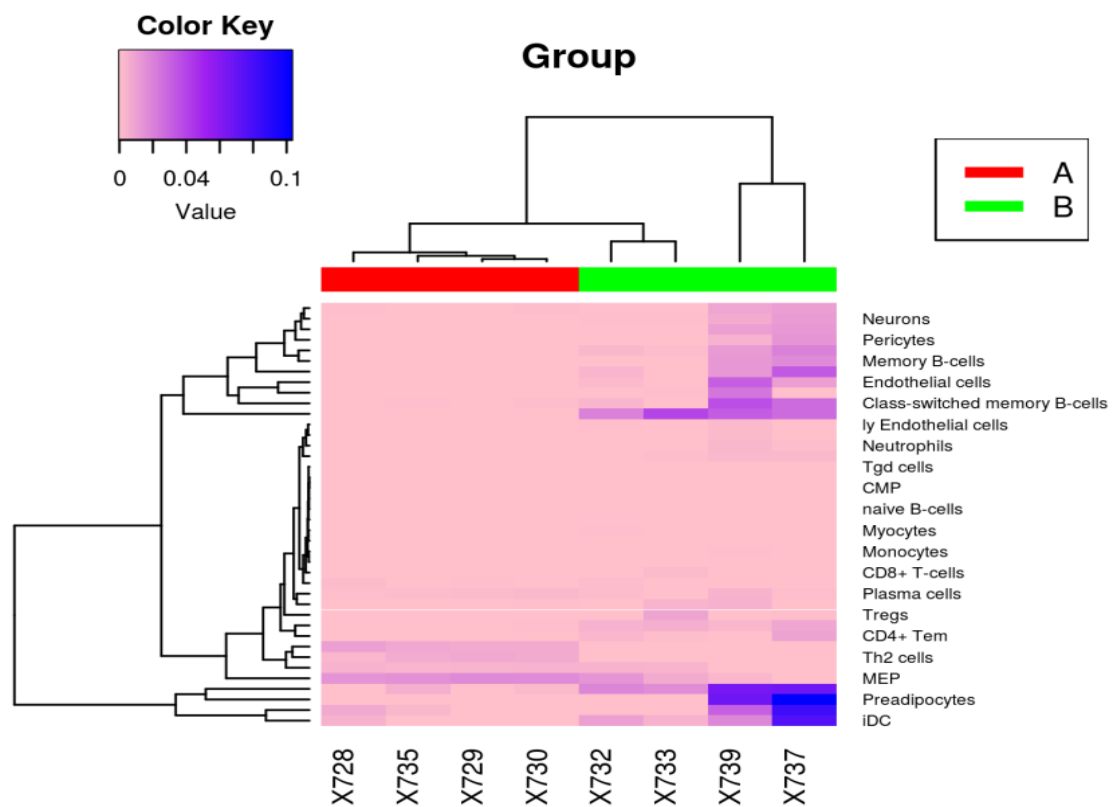
3.3.7 ΨΗΦΙΑΚΗ ΚΥΤΤΑΡΟΜΕΤΡΙΑ ΡΟΗΣ

• CIBERSORT-X

Έγινε προσπάθεια να γίνει προσδιορισμός των κυτταρικών πληθυσμών σε κάθε δείγμα, με τη βοήθεια του προγράμματος CIBERSORT-X. Ωστόσο, τα αποτελέσματα που προέκυψαν δεν ήταν στατιστικά σημαντικά (τιμή $p > 0.05$).

• X-CELL

Το πρόγραμμα xCell χρησιμοποιήθηκε στα δείγματα μας, και έπειτα με τη χρήση της γλώσσας προγραμματισμού R, οικοδομήθηκε ένα heatmap. Τα αποτελέσματα που προέκυψαν δεν ήταν στατιστικά σημαντικά (τιμή $p > 0.05$) (Εικόνα 26).



Εικόνα 26. Heatmap που προέκυψε από το πρόγραμμα Xcell

ΚΕΦΑΛΑΙΟ 4 - ΣΥΖΗΤΗΣΗ

Ο λόγος, που οι περισσότεροι άνθρωποι με νόσο COVID-19 έχουν σχετικά ήπια συμπτώματα, ενώ άλλοι αναπτύσσουν αναπνευστική ανεπάρκεια, με απειλητικές για τη ζωή επιπλοκές, παραμένει άγνωστος. Τα δεδομένα δείχνουν ότι η κλινική εικόνα εξαρτάται σε μεγάλο βαθμό από την ανοσολογική απόκριση (Bibert et al., 2021). Η ιδανική ανοσολογική απόκριση στον SARS-CoV-2 χαρακτηρίζεται από αυξημένο επιθηλιακό φραγμό, ήπια φλεγμονή και ισχυρή CD8+ T κυτταρική απόκριση (Pan et al., 2021). Η ανάλυση του μεταγραφώματος ρινοφαρυγγικών δειγμάτων μπορεί να είναι ένα ισχυρό εργαλείο ποσοτικοποίησης της ανοσολογικής απόκρισης του ξενιστή και μπορεί να παρέχει πολύτιμη βοήθεια στη διαλεύκανση της παθοφυσιολογίας της νόσου COVID-19 (Jain et al., 2021).

Στις ιογενείς λοιμώξεις, το ιικό φορτίο θεωρείται προγνωστικός δείκτης της σοβαρότητας και της εξέλιξης της νόσου (Fajnzylber et al., 2020). Ο SARS-CoV-2 δεν αποτελεί εξαίρεση. Οι ασθενείς που έχουν διαγνωστεί με COVID-19 μπορεί να έχουν ανιχνεύσιμο SARS-CoV-2 RNA σε δείγματα από την ανώτερη αναπνευστική οδό για εβδομάδες μετά την εμφάνιση των συμπτωμάτων. Ωστόσο, η παρατεταμένη ή η επαναλαμβανόμενη ανίχνευση ιικού RNA δεν υποδηλώνει απαραίτητα παρατεταμένη μολυσματικότητα. Μολυσματικότητα σε ασθενείς με θετικό RT-PCR SARS-CoV-2 test πάνω από 10 ημέρες μετά την έναρξη ήπιας νόσησης έχει σπάνια τεκμηριωθεί. Επιπλέον, τα επίπεδα του ιικού RNA, κατά τη διάρκεια της φάσης της ανάρρωσης από νόσο COVID-19, εάν είναι ανιχνεύσιμα, είναι χαμηλότερα από αυτά που ανιχνεύονται κατά την οξεία φάση της λοίμωξης. Το Κέντρο Ελέγχου και Πρόληψης Νοσημάτων (Centers for Disease Control and Prevention, CDC) υποστηρίζει ότι ένα θετικό τεστ εντός τριών μηνών μετά τη νόσηση από COVID-19, ενέχει χαμηλή πιθανότητα ενεργής λοίμωξης. Σε μία άλλη μελέτη διαπιστώθηκε ότι ο μέσος χρόνος αρνητικοποίησης του RT-PCR test ήταν 24 ημέρες μετά την εμφάνιση των συμπτωμάτων και το 10 % παρέμεινε θετικό ακόμη και 33 ημέρες μετά την έναρξη των συμπτωμάτων (Mandal et al., 2020).

Σε μία μελέτη που δημοσιεύτηκε στο Lancet, η παρατεταμένη θετικότητα του RT-PCR SARS-CoV-2 test, σε άτομα που έχουν νοσήσει με COVID-19, έχει συσχετιστεί με αυξημένη κυτταρική ανοσοαπόκριση (Gombar et al., 2020). Έπειτα, σύμφωνα με μία άλλη μελέτη, στο πλάσμα των βαρέως πασχόντων ασθενών, συγκεκριμένα των μη επιζώντων, παρατηρείται υψηλότερο ιικό φορτίο RNA σε σύγκριση με τις άλλες ομάδες ασθενών (Bermejo-Martin et al., 2020). Επίσης, έχει διαπιστωθεί ότι το ιικό φορτίο είναι υψηλότερο στους ασθενείς, που νοσούν πιο βαριά, υποδηλώνοντας ότι το ιικό φορτίο μπορεί να συσχετίζεται με τη θνησιμότητα (Bergamaschi et al., 2021). Επιπλέον, υποστηρίζεται ότι το ιικό φορτίο SARS-CoV-2 συσχετίζεται ανεξάρτητα και προβλέπει τη θνησιμότητα (Pujadas et al., 2020). Εν αντιθέσει, η μακροχρόνια μόλυνση με SARS-CoV-2 χαρακτηρίζεται από χαμηλό ιικό φορτίο και χαμηλή παθογένεια από τον ιό (X. Wang et al., 2020).

Σε αυτό το πλαίσιο, η μελέτη μας διερεύνησε εάν υπάρχει διαφορά στην αρχική ανοσολογική απόκριση μεταξύ ασθενών με COVID-19 με διαφορετικό χρόνο αρνητικοποίησης του PCR test SARS-CoV-2, κάτι το οποίο σύμφωνα με τα άνωθεν, μπορεί να συσχετίζεται με τη σοβαρότητα της νόσου. Επιπρόσθετα, για το σκοπό αυτό, αναπτύχθηκε βιοπληροφορική ροή ανάλυσης δεδομένων QuantSeq, με χρήση και αξιοποίηση διαφόρων εργαλείων με στόχο την ποιοτική τους σύγκριση.

Αναφορικά με τη σύγκριση των αποτελεσμάτων μας με τη βιβλιογραφία, υπήρξε ο περιορισμός ότι έπειτα από αρκετά λεπτομερή αναζήτηση των βάσεων δεδομένων ευρείας εμβέλειας (PUBMED, GOOGLE SCHOLAR), για εύρεση παρόμοιου ερευνητικού ερωτήματος, δεν βρέθηκε κάποια αντίστοιχη μελέτη.

4.1 ΑΝΑΛΥΣΗ ΔΙΑΦΟΡΙΚΗΣ ΕΚΦΡΑΣΗΣ

Στη βιβλιογραφία, έχει αναφερθεί ότι DEGs που συσχετίζονται με την απορρύθμιση της πήξης του αίματος και της ομοιόστασης, συνδέονται με την πρόοδο της νόσου. Επιπρόσθετα, έχει σημειωθεί ότι DEGs, που σχετίζονται

με την ενεργοποίηση της ανοσολογικής απόκρισης του ξενιστή, τη σηματοδότηση των κυτταροκινών, την ενδοκυττάρωση ιών είναι αυξημένα σε σοβαρές περιπτώσεις COVID-19, σε σχέση με μέτριες περιπτώσεις (Madewell et al., 2020). Ανάλυση διαφορικής έκφρασης σε ρινοφαρυγγικά δείγματα ασθενών με COVID-19 σε σχέση με υγιείς έχει δείξει έντονη ενεργοποίηση γονιδίων που σχετίζονται με σηματοδότηση IFN και ISGs, αλλά καταστολή των γονιδίων που σχετίζονται με την IL-6 και την IL-8 (Ng et al., 2021). Η έκφραση των προφλεγμονωδών κυτταροκινών και χημειοκινών είναι απαραίτητες για τον αρχικό καθαρισμό της ιικής λοίμωξης, αν και η υπερβολική έκφραση τους έχει συνδεθεί με το ARDS. Ο SARS-CoV2 μπορεί να έχει ολέθριες συνέπειες για τον οργανισμό, πυροδοτώντας καταρράκτη κυτταροκινών (Costela-ruiz et al., 2020). Παράλληλα, σε άλλη μελέτη βρέθηκε ότι σε ασθενείς με COVID-19, παρατηρείται υπερέκφραση των γονιδίων που συσχετίζονται με την δραστηριότητα των κυτταροκινών και με το μονοπάτι σηματοδότησης των ιντερφερονών, καθώς και υποέκφραση των γονιδίων που εμπλέκονται στην οξειδωτική φωσφορυλίωση και την ιική μεταγραφή (Vilas et al., 2020). Η IFN γ , ή ιντερφερόνη τύπου II, είναι μια κυτταροκίνη που είναι κρίσιμη για την ειδική και μη ειδική ανοσία έναντι των ιογενών λοιμώξεων. Η IFN γ είναι σημαντικός ενεργοποιητής μακροφάγων και επαγωγέας έκφρασης μορίων μείζονος συμπλόκου ιστοσυμβατότητας κατηγορίας II. Η σημασία της IFN γ στο ανοσοποιητικό σύστημα πηγάζει εν μέρει από την ικανότητά της να αναστέλλει τον πολλαπλασιασμό του ιού απευθείας, και κυρίως από τις ανοσοδιεγερτικές και ανοσορρυθμιστικές της επιδράσεις. Η IFN γ παράγεται κυρίως από κύτταρα-φυσικούς φονείς (NK) ως μέρος της μη ειδικής ανοσολογικής απόκρισης, αλλά και από τα CD4 Th1 και CD8 κυτταροτοξικά κύτταρα T λεμφοκύτταρα, εφόσον ενεργοποιηθεί η ειδική ανοσία. Έχει βρεθεί ότι γονίδια που συσχετίζονται με τη σηματοδότηση των ιντερφερονών και δη της IFN γ , είναι στατιστικά σημαντικά αυξημένα σε πνεύμονα πιθήκων macaque, που νοσούν με COVID-19 (Rosa et al., 2021). Επίσης, έχει επισημανθεί στη βιβλιογραφία ότι τα επίπεδα της IFN γ τείνουν να είναι μικρότερα σε ασθενείς που νοσούν βαριά με COVID-19, σε σύγκριση με ασθενείς που νοσούν πιο ήπια (Awan, 2021).

Στη λίστα με τα διαφορεικά εκφραζόμενα γονίδια, σημαντικός αριθμός γονιδίων ήταν IFN ρυθμιζόμενα. Παράδειγμα είναι πρωτεΐνες του ριβοσώματος (πχ RPL11P3, RPL18AP15). Οι πρωτεΐνες RP αξιοποιούνται από διάφορους ιούς, συμπεριλαμβανομένου του SARS-CoV-2, κατά τη διάρκεια της μόλυνσης για να τερματιστεί η μετάφραση των πρωτεϊνών του ξενιστή και να διευκολυνθεί η μετάφραση των ιικών πρωτεϊνών (Biji et al., 2021). Οι πρωτεΐνες MAP3K7 και IL10RB στη μελέτη μας βρέθηκαν υπερεκφρασμένες στην ομάδα A, κάτι που μπορεί να αποτελεί δείκτη της μεγαλύτερης σοβαρότητας της λοίμωξης στην ομάδα A.

Στη μελέτη μας, δεν βρέθηκε στατιστικά σημαντική διαφορά στην έκφραση των γονιδίων ACE2, TMPRSS2, BSG, CTSL, DPP4, τα οποία σύμφωνα με τους Sungnak και συναδέλφους, έχουν επισημανθεί να συσχετίζονται με τη είσοδο του ιού στα κύτταρα του ξενιστή (Sungnak et al., 2020). Βρέθηκε στατιστικά σημαντική διαφορά στην έκφραση του TLR3, ο οποίος είναι ένας υποδοχέας αναγνώρισης προτύπου (Pattern Recognition Receptor, PRR) και παίζει κομβικό ρόλο στην ενεργοποίηση της μη ειδικής ανοσίας (Warren, 2005). Η ενεργοποίηση των μονοπατιών των TLRs οδηγεί στην έκκριση προφλεγμονωδών κυτταροκινών, όπως οι IL-1, IL-6, TNF- α (Tumor Necrosis Factor- α), IFN-I (Khanmohammadi & Rezaei, 2021). Αναφορικά με τις ιντεγκρίνες, βρέθηκε στατιστικά σημαντική διαφορά στη γονιδιακή έκφραση των κάτωθι γονιδίων: ITGA9, ITGAX, ITGA8, ITGA11, ITGA2B, ITGA3, ITGAM. Προηγούμενες αναφορές, έχουν επισημάνει ότι οι ιντεγκρίνες συμμετέχουν στην παθογένεση του SARS-CoV-2 (Islam et al., 2021). Ειδικότερα, η πρωτεΐνη-ακίδα του SARS-CoV-2 μπορεί να συνδεθεί με τις ιντεγκρίνες και έτσι στους ασθενείς που υπάρχει μεγαλύτερη έκφραση ιντεγκρινών, μπορούμε να υποθέσουμε ότι ο SARS-CoV-2 μπορεί να διεισδύσει πιο εύκολα στα κύτταρα. Επίσης, δεν βρέθηκε διαφορά στην έκφραση των γονιδίων, NF- κ B, TBK1, τα οποία εμπλέκονται στην ενεργοποίηση της IFN-I. Σε άλλες μελέτες έχει βρεθεί ότι η έκφραση των προφλεγμονωδών κυτταροκινών υποκινείται από την πρωτεΐνη NF- κ B (Su et al., 2021).

Σε άλλη μελέτη, που έγινε ανάλυση του ολικού μεταγραφώματος σε ρινοφαρυγγικά δείγματα ασθενών με COVID-19, βρέθηκε αύξηση στην έκφραση υποδοχέων IRs (Immune Inhibitory Receptors) στο ανώτερο αναπνευστικό των ασθενών, καθώς και των δεικτών των T κυττάρων (CD25 και CD69) και των μακροφάγων (CD86 και HLA-DRA) (Saheb Sharif-Askari et al., 2021). Από τα IRs γονίδια, στη μελέτη μας βρέθηκε διαφορά στην έκφραση μεταξύ των δύο ομάδων του γονιδίου HLA-DRA. Η πρωτεΐνη HLA-DRA διαδραματίζει σημαντικό ρόλο στον ανοσοποιητικό σύστημα, δεδομένου ότι διευκολύνει την αντιγονοπαρουσίαση. Επιπρόσθετα, στην ίδια μελέτη βρέθηκε ότι η έκφραση των εξής γονιδίων BTLA, LAG3, FCGR2B, PDCD1, CEACAM1, CTLA4, CD72, SIGLEC7 συσχετίζεται με το ιικό φορτίο. Στη δικιά μας μελέτη, βρέθηκε ότι μόνο το γονίδιο SIGLEC7 από τα άνωθεν είναι υπερεκφρασμένο στην ομάδα A.

Κατόπιν, το γονίδιο της NRP2, το οποίο είναι υπερεκφρασμένο στην ομάδα A, σε σχέση με την ομάδα B, έχει βρεθεί στη βιβλιογραφία να συσχετίζεται με το κατά πόσο μία λοίμωξη είναι συμπτωματική ή όχι (Qi et al., 2021). Επίσης, το γονίδιο NDFIP2, το οποίο συνδέεται με την αποκοκκίωση των ουδετερόφιλων και τη NF-κB σηματοδότηση, στη μελέτη μας ήταν υπερεκφρασμένο στην ομάδα A, και αυτή η παρατήρηση συμβαδίζει με τη βιβλιογραφία (Qi et al., 2021).

Έπειτα, αυξημένη έκφραση ειδικών μονοπατιών γονιδίων ιντερφερονών (SHFL, IFI6, IFI27, και IFIT1), που έχουν προηγούμενα συσχετιστεί με τη μη ειδική αντιική ανοσία έναντι άλλων ιών, έχουν παρατηρηθεί σε άλλες μελέτες (Butler et al., 2021). Ένας άλλος παράγοντας, ο LCN2, καθώς και πρωτεάσες τύπου SERPIN, που έχουν διαπιστωθεί να επηρεάζουν την πρόγνωση των ασθενών με COVID-19, στη μελέτη μας δεν είχαν διαφορετική έκφραση μεταξύ των δύο ομάδων (Jean-F. Tomb et al., 2020). Τα γονίδια ANXA1 και FPR1 που βρέθηκαν στατιστικά σημαντικά διαφορετικά εκφρασμένα στις δύο ομάδες, έχει βρεθεί να διαφέρουν μεταξύ μετρίως νοσούντων και σοβαρών νοσούντων με COVID-19 (Ren et al., 2021).

Επιπροσθέτως, στη μελέτη μας, βρέθηκαν διαφορεικά εκφρασμένα γονίδια τα S100A2 και S100A14. Αυτές οι πρωτεΐνες συμμετέχουν στη ρύθμιση της ανοσολογικής απάντησης. Ενεργοποιούν τη σηματοδότηση NF-κB, διεγείροντας τα κύτταρα να παράγουν προφλεγμονώδεις κυτταροκίνες στο σημείο της λοίμωξης (Ren et al., 2021). Σε μία μετα ανάλυση έχει βρεθεί ότι γονίδια της οικογένειας S100, συσχετίζονται με την πρόγνωση της νόσου COVID-19, και ειδικότερα, είναι προγνωστικοί δείκτες της σοβαρής νόσου COVID-19 (Biji et al., 2021).

Το γονίδιο IFITM3, το οποίο έχει βρεθεί σε άλλη μελέτη υπερεκφρασμένο σε ασθενείς με COVID-19 σε σχέση με υγιείς, βρέθηκε υπερεκφρασμένο στην ομάδα A (Melo et al., 2020).

Αναφορικά με τις κασπάσες, τα κάτωθι γονίδια CARD14 (caspase recruitment domain family member 14), CASP16P (caspase 16), CARD18 (caspase recruitment domain family member 18) βρέθηκαν διαφορετικά εκφρασμένα μεταξύ των δύο ομάδων. Σε μία άλλη μελέτη έχει προταθεί ότι η CARD-14, μπορεί να διαδραματίζει σημαντικό ρόλο στην εξέλιξη της νόσου COVID-19 σε σοβαρή (Kelleni, 2021). Η αναγνώριση του ιού από τα κύτταρα του ξενιστή έχει ως αποτέλεσμα την ενεργοποίηση των κασπασών. Έχει υποτεθεί ότι μία υπερβολική ενεργοποίηση των κασπασών στη νόσο COVID-19 μπορεί να οδηγήσει σε σοβαρές επιπλοκές (Vaziri et al., 2020).

4.2 ΛΕΙΤΟΥΡΓΙΚΗ ΑΝΑΛΥΣΗ

Στη βιβλιογραφία έχει αναφερθεί ότι τα μονοπάτια της μη ειδικής ανοσίας και της φλεγμονής ενεργοποιούνται με τρόπο που εξαρτάται από τη σοβαρότητα της νόσησης με COVID-19. Μονοπάτια που εμπλουτίζονται σε λοίμωξη με SARS-CoV-2, περιλαμβάνουν μονοπάτια σχετικά με την μη ειδική ανοσία, τη σηματοδότηση ISGs, την ωρίμανση των δενδριτικών κυττάρων, τη σηματοδότηση μέσω της IL-2, τη σηματοδότηση μέσω της IFN γ και τη MAPK σηματοδότησης (Ng et al., 2021, Mun et al., 2021). Άλλοι όροι που έχουν βρεθεί εμπλουτισμένοι, σε λειτουργική ανάλυση σε ρινοφαρυγγικά δείγματα,

είναι οι κάτωθι: ιικές διαδικασίες (viral process), ανοσολογική απόκριση (immune response), μη ειδική ανοσολογική απόκριση (innate immune response), άμυνα εναντίον ιών (defense response to virus), σηματοδότηση ιντερφερονών (interferon signaling) (Islam et al., 2021). Άλλη μελέτη έχει καταλήξει ότι σε ρινοφαρυγγικά δείγματα σε ασθενείς με COVID-19, ευρίσκονται εμπλουτισμένα μονοπάτια που σχετίζονται με την αυτοφαγία/μιτοχονδριακές διαδικασίες (Singh et al., 2021).

Στη μελέτη μας βρέθηκε εμπλουτισμένη η γονιδιακή οντολογία interferon-gamma-mediated signaling pathway. Στη βιβλιογραφία έχει σημειωθεί ο ρόλος της IFN γ στην εξέλιξη της λοίμωξης. Η IFN γ στο αίμα έχει βρεθεί να είναι υψηλότερη σε ασθενείς που πέθαναν, παρά σε αυτούς που επέζησαν, ενώ επίσης η IFN γ ήταν υψηλότερη στα πρώιμα στάδια της νόσου (Carolina et al., 2020). Επίσης, σε μία άλλη μελέτη, έχει επισημανθεί ότι χαμηλότερη κυκλοφορούσα IFN γ είναι παράγοντας κινδύνου για πνευμονική ίνωση σε ασθενείς με COVID-19. Επίσης, έχει διερευνηθεί η θετική επίδραση της IFN γ στην κλινικής σταθεροποίηση και ανάρρωση ασθενών με πνευμονία από COVID-19 (Al et al., 2021). Νοσηλευόμενοι και εξωτερικοί ασθενείς παρουσίασαν αυξημένη ρύθμιση των οδών που σχετίζονται με την ιντερφερόνη, αν και παρατηρήθηκαν αυξημένες και πιο ισχυρές φλεγμονώδεις ανπιδράσεις σε νοσηλευόμενους ασθενείς με πιο κλινικά σοβαρή ασθένεια (Ng et al., 2021). Η απόκριση σε ιντερφερόνη I και II είναι πιο ισχυρή σε ασθενείς που νοσούν με μέτρια βαρύτητα, σε σύγκριση με ασθενείς που νοσούν με ήπια ή σοβαρή βαρύτητα (Y et al., 2021, Zhang et al., 2021). Τα επίπεδα της IFN γ είναι αυξημένα κυρίως σε ήπια ή μετρίως νοσούντες ασθενείς και ελαφρά αυξημένα σε σοβαρά νοσούντες ασθενείς, αλλά όχι σε αυτούς σε κρίσιμη κατάσταση. Στην κορύφωση της νόσου, τα επίπεδα των ιντερφερονών είναι αυξημένα, ενώ σε ασθενείς που είναι σε κρίσιμη κατάσταση, μετάγραφα ρυθμιστικά της απόκρισης τύπου I, όπως IRF3/7/8 και MAVS καταστέλλονται (Zhang et al., 2021).

Σε ένα άλλο άρθρο, περιγράφηκε ότι το μεταγράφομα της λοίμωξης με SARS-CoV-2, αποτελείται από 866 γονίδια, που συμμετέχουν κυρίως στις

λειτουργίες του ανοσοποιητικού συστήματος: ενεργοποίηση λευκοκυττάρων, σηματοδότηση κυτταροκινών, απόκριση σε ερεθίσματα. Επίσης, βρέθηκαν υπερεκφρασμένα γονίδια που σχετίζονται με τη δραστηριότητα ιόντων, κάτι το οποίο βρέθηκε και στα δικά μας αποτελέσματα, δεδομένου από τη λειτουργική ανάλυση βρήκαμε εμπλουτισμό του όρου calcium signalling. Λιγότερα γονίδια, βρέθηκαν να συσχετίζονται με την ιική λοίμωξη, όπως η κυτταρική διαφοροποίηση, η διαπερατότητα του ενδοθηλίου, η κυτταρική αύξηση (Nour et al., 2021).

Στη μελέτη μας βρέθηκε εμπλουτισμένο το μονοπάτι cell adhesion, κάτι το οποίο έχει βρεθεί και στη μελέτη των Gu και συναδέλφων- βρέθηκε εμπλουτισμένος ο όρος “cadherin binding involved in cell-cell adhesion”(Gu & Yuan, 2020). Επίσης, σε μία μελέτη που εξέτασαν BALF δείγματα, βρέθηκε στατιστικά σημαντικός εμπλουτισμός όρων, που σχετίζονται με το cell adhesion, όπως cell–substrate adhesion, adherens junction organization, focal adhesion assembly and adherens junction assembly. (Li et al., 2020). Τα λευκοκύτταρα αλληλεπιδρούν με τα μόρια κυτταρικής προσκόλλησης του ενδοθηλίου (endothelial cell adhesion molecules, CAM) και κατόπιν μεταναστεύουν διαμέσω του ενδοθηλίου.

4.3. ΣΥΓΚΡΙΣΗ ΒΙΟΠΛΗΡΟΦΟΡΙΚΩΝ ΡΟΩΝ

Το RNA-SEQ είναι πλέον η τεχνολογία επιλογής για πειράματα διαφορικής γονιδιακής έκφρασης, αλλά δεν είναι σαφές ποια στατιστικά εργαλεία είναι καλύτερα για την ανάλυση των δεδομένων (Schurch et al., 2016). Υπάρχει έλλειψη ομοφωνίας για τη βέλτιστη κατασκευή μίας βιοπληροφορικής ροής για την ανάλυση RNA-SEQ (C. D. Zhang et al., 2018). Σύμφωνα με τη βιβλιογραφία, δεν παρατηρείται στατιστικά σημαντική διαφορά στα αποτελέσματα, ανάλογα με τον αλγόριθμο trimming και στοίχισης (alignment) που χρησιμοποιείται (Schaarschmidt et al., 2020, Corchete et al., 2020). Αναφορικά με την ψευδοστοίχιση, έχει δείχθει ότι ένα από τα κύρια πλεονεκτήματα των αλγορίθμων που την χρησιμοποιούν είναι ότι μπορούν να εκτελέσουν τις διαδικασίες της στοίχισης, της ποσοτικοποίησης και της

κανονικοποίησης σε ένα βήμα. Ροές βιοπληροφορικής που βασίζονται στην ψευδοστοίχιση, έχουν ικανοποιητική ακρίβεια (precision) στη διαφορική έκφραση γονιδίων, αλλά η ορθότητα (accuracy) είναι κατώτερη των κλασικών εργαλείων στοίχισης (aligners). Σε κάθε περίπτωση, οι ψευδοστοιχιστές είναι μία καλή εναλλακτική στη διερεύνηση των δεδομένων του RNA-SEQ, δεδομένου ότι ο χρόνος εκτέλεσης είναι πολύ μικρότερος από τις συμβατικές μεθόδους στοίχισης. Μία μελέτη, που συνέκρινε την επίδοση διαφορετικών στοιχιστών στην επεξεργασία δεδομένων RNA-SEQ, κατέληξε ότι το QuantSeq, σε συνδυασμό με το Salmon, μπορεί να έχει ικανοποιητικά αποτελέσματα στην ποσοτικοποίηση της γονιδιακής έκφρασης (Corley et al., 2019).

Απεναντίας, το βήμα της κανονικοποίησης είναι κρίσιμο για την επεξεργασία των δεδομένων του RNA-SEQ (Corchete et al., 2020). Στην ίδια μελέτη αποδείχθηκε ότι οι αλγόριθμοι κανονικοποίησης που χρησιμοποιούν τη μέθοδο TMM (Trimmed Mean of the M-values) είναι ανώτεροι των μεθόδων που χρησιμοποιούν τις μεθόδους RPKM (Reads Per Kilobase Million), FPKM (Fragments Per Kilobase Million), TPM (Transcripts Per Kilobase Million) ή λοιπές μεθόδους. Το DESeq2 και το edgeR είναι πολύ παρόμοια και αμφότερα υποθέτουν ότι κανένα γονίδιο δεν εκφράζεται διαφορετικά. Το DESeq2 χρησιμοποιεί μια "γεωμετρική" στρατηγική κανονικοποίησης, ενώ το edgeR είναι μία μέθοδος, που βασίζεται σε σταθμισμένο μέσο λογαριθμικών αναλογιών. Και οι δύο αλγόριθμοι κανονικοποιούν τα δεδομένα αρχικά μέσω του υπολογισμού του μεγέθους / των παραγόντων κανονικοποίησης. Έχει βρεθεί ότι, το DESeq2 και το edgeR έχουν παρόμοια αποτελεσματικότητα.

ΚΕΦΑΛΑΙΟ 5 - ΣΥΜΠΕΡΑΣΜΑ – ΠΕΡΙΟΡΙΣΜΟΙ – ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΤΟ ΜΕΛΛΟΝ

5.1 ΣΥΜΠΕΡΑΣΜΑ

Συμπερασματικά, στη συγκεκριμένη εργασία, πραγματοποιήσαμε ένα πείραμα αλληλούχισης, και συγκεκριμένα QuantSeq, σε ρινοφαρυγγικά δείγματα από ασθενείς με COVID-19. Συγκρίναμε την αρχική ανοσολογική απόκριση μεταξύ ασθενών που είχαν θετικό RT-PCR SARS-CoV-2 test ένα μήνα μετά την αρχική λοίμωξη (θετικοποίηση του πρώτου RT-PCR SARS-CoV-2 test), έναντι ασθενών που είχαν αρνητικό το δεύτερο test. Για το σκοπό αυτό, αρχικά εκτελέσαμε το εργαστηριακό μέρος της απομόνωσης RNA και της αλληλούχισης, με το πρωτόκολλο QuantSeq 3' mRNA-SEQ Library Prep Kit FWD για Illumina (Lexogen), αναγνώσεις μονής ανάγνωσης (single-end reads). Κατόπιν, προχωρήσαμε στην ανάπτυξη βιοπληροφορικών ροών με σκοπό τη διαφορική έκφραση και τη λειτουργική ανάλυση των γονιδίων, που προέκυψαν. Το QuantSeq είναι μία ικανοποιητική εναλλακτική στο total RNA-SEQ, το οποίο επιτρέπει την αξιόπιστη ποσοτικοποίηση της γονιδιακής έκφρασης, απαιτώντας λιγότερο χρόνο, μικρότερη ποιότητα και ποσότητα δειγμάτων προς ανάλυση και πιο εύκολη διαδικασία. Για τη στοίχιση χρησιμοποιήσαμε τα εργαλεία STAR και SALMON, για τη διαφορική έκφραση τα εργαλεία DESeq2 και edgeR και για τη λειτουργική ανάλυση τα εργαλεία DAVID, g:Profiler. Αναφορικά με τη βέλτιστη βιοπληροφορική ροή με βάση την ποιοτική σύγκριση, καταλήξαμε ότι επρόκειτο για την κάτωθι: STAR \Rightarrow DESeq2 \Rightarrow DAVID. Από την ανάλυση διαφορικής έκφρασης βρήκαμε 4464 διαφορικά εκφρασμένα γονίδια σε επίπεδο σημαντικότητας 5%, $|\log_2\text{foldchange}| > 1$. Επίσης, από τη λειτουργική ανάλυση, βρήκαμε ότι οι γονιδιακές οντολογίες που εμπλουτίστηκαν στατιστικά σημαντικά ήταν οι: cell adhesion, chemical synaptic transmission, interferon-gamma-mediated και τα μονοπάτια: Calcium signaling pathway, Morphine addiction, Axon guidance, Cell adhesion molecules (CAMs), Osteoclast differentiation. Έτσι, μπορούμε να υποθέσουμε ότι παρατηρείται διαφορά στην αρχική ανοσολογική απόκριση

των ασθενών με COVID-19, που εξαρτάται από το χρόνο αρνητικοποίησης του RT-PCR SARS-CoV-2 test, καθώς και ότι η βιοπληροφορική ροή ανάλυσης δεδομένων 3'QuantSeq που αναπτύξαμε είναι ικανοποιητική για τη μελέτη της ανοσοβιολογικής απόκρισης στον SARS-CoV-2.

5.2 ΠΕΡΙΟΡΙΣΜΟΙ – ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΤΟ ΜΕΛΛΟΝ

Παρότι η μελέτη μας ήταν καινοτόμα, δεδομένου ότι διερεύνησε ένα ερώτημα, που δεν έχει εξεταστεί ενδελεχώς - εάν η παρατεταμένη θετικότητα του RT-PCR test συσχετίζεται με διαφορά στην αρχική ανοσολογική απόκριση- και έκανε χρήση μίας πρωτοποριακής μεθοδολογίας (3'QuantSeq) και ανάπτυξη εξειδικευμένης βιοπληροφορικής ανάλυσης, είχε αρκετούς περιορισμούς, που στο μέλλον καλό θα ήταν να ληφθούν υπόψιν. Ειδικότερα, δεν υπήρχαν πολλά κλινικά δεδομένα για τα δείγματα, και για αυτό δεν αξιοποιήθηκαν στην ανάλυση. Λόγω της έλλειψης κλινικών πληροφοριών, η ερμηνεία των αποτελεσμάτων είναι υπό αμφισβήτηση. Επιπρόσθετα, δεν υπήρξε follow-up των ασθενών και μελέτη της ανοσολογικής απόκρισης των σε μεταγενέστερο χρόνο από την πρωτοεμφάνιση της λοίμωξης. Ακόμη, το δείγμα της μελέτης ήταν μικρό. Εκτός αυτών, σχετικά με τη σύγκριση των εργαλείων που χρησιμοποιήθηκαν στις βιοπληροφορικές ροές, εκτός από την απλή ποιοτική σύγκριση, μία στατιστικά τεκμηριωμένη σύγκριση των βιοπληροφορικών ροών κρίνεται ωφέλιμη και αναγκαία για την εξαγωγή ενός πιο αξιόπιστου συμπεράσματος.

ΚΕΦΑΛΑΙΟ 6 - ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ

6.1 ΑΓΓΛΟΦΩΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Al, M., Sa, B., Pa, T., & Fl, E. (2021). [Interferon gamma in the treatment of patients with moderate COVID-19]. *Voprosy Virusologii*, 66(1), 47–54. <https://doi.org/10.36233/0507-4088-24>
2. Alfego, D., Sullivan, A., Poirier, B., Williams, J., Adcock, D., & Letovsky, S. (2021). A population-based analysis of the longevity of SARS-CoV-2 antibody seropositivity in the United States. *EClinicalMedicine* 100902. <https://doi.org/10.1016/j.eclinm.2021.100902>
3. Aran, D., Hu, Z., & Butte, A. J. (2017). xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biology*, 18(1), 1–14. <https://doi.org/10.1186/s13059-017-1349-1>
4. Awan, A. (2021). GAPDH, Interferon γ , and Nitric Oxide: Inhibitors of Coronaviruses. *Frontiers in Virology*, 0, 18. <https://doi.org/10.3389/FVIRO.2021.682136>
5. *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*. (n.d.). Retrieved October 21, 2021, from <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
6. *Bash - GNU Project - Free Software Foundation*. (n.d.). Retrieved October 21, 2021, from <https://www.gnu.org/software/bash/>
7. *BBDuk Guide - DOE Joint Genome Institute*. (n.d.). Retrieved October 24, 2021, from <https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbduk-guide/>
8. Bertholee, D., Maring, J. G., & van Kuilenburg, A. B. P. (2017). Genotypes Affecting the Pharmacokinetics of Anticancer Drugs. *Clinical Pharmacokinetics*, 56(4), 317–337. <https://doi.org/10.1007/s40262-016-0450-z>
9. Bibert, S., Guex, N., Lourenco, J., Brahier, T., Papadimitriou-Olivgeris, M., Damonti, L., Manuel, O., Liechti, R., Götz, L., Tschopp, J., Quinodoz, M., Vollenweider, P., Pagani, J. L., Oddo, M., Hügli, O., Lamoth, F., Erard, V., Voide, C., Delorenzi, M., ... Bochud, P. Y.

- (2021). Transcriptomic Signature Differences Between SARS-CoV-2 and Influenza Virus Infected Patients. *Frontiers in Immunology*, 12(May), 1–25. <https://doi.org/10.3389/fimmu.2021.666163>
10. Biji, A., Khatun, O., Swaraj, S., Narayan, R., Rajmani, R. S., Sardar, R., Satish, D., Mehta, S., Bindhu, H., Jeevan, M., Saini, D. K., Singh, A., Gupta, D., & Tripathi, S. (2021). Identification of COVID-19 prognostic markers and therapeutic targets through meta-analysis and validation of Omics data from nasopharyngeal samples. *EBioMedicine*, 70, 103525. <https://doi.org/10.1016/j.ebiom.2021.103525>
11. *Bioconductor - About*. (n.d.). Retrieved October 21, 2021, from <https://www.bioconductor.org/about/>
12. Butler, D., Mozsary, C., Meydan, C., Foox, J., Rosiene, J., Shaiber, A., Danko, D., Afshinnkoo, E., MacKay, M., Sedlazeck, F. J., Ivanov, N. A., Sierra, M., Pohle, D., Zietz, M., Gisladdottir, U., Ramlall, V., Sholle, E. T., Schenck, E. J., Westover, C. D., Mason, C. E. (2021). Shotgun transcriptome, spatial omics, and isothermal profiling of SARS-CoV-2 infection reveals unique host responses, viral diversification, and drug interactions. *Nature Communications*, 12(1), 1–17. <https://doi.org/10.1038/s41467-021-21361-7>
13. Carolina, A., Castro, M. De, Paulo, J., Wind, R., Goes, M., Garcia, R., Ossoski, C., Michalski, A., Padua, D., Noronha, L. De, Moreno-amaral, A., Pellegrino, C., & Francisco, F. (2020). *Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information . January.*
14. Choi, H., & Shin, E. C. (2021). Roles of type i and iii interferons in covid-19. *Yonsei Medical Journal*, 62(5), 381–390. <https://doi.org/10.3349/ymj.2021.62.5.381>
15. Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2009). The Sanger FASTQ file format for sequences with quality

- scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38(6), 1767–1771. <https://doi.org/10.1093/nar/gkp1137>
16. Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1), 1–19. <https://doi.org/10.1186/s13059-016-0881-8>
17. Consortium, T. G. O. (2008). The Gene Ontology project in 2008. *Nucleic Acids Research*, 36(suppl_1), D440–D444. <https://doi.org/10.1093/NAR/GKM883>
18. Corchete, L. A., Rojas, E. A., Alonso-López, D., De Las Rivas, J., Gutiérrez, N. C., & Burguillo, F. J. (2020). Systematic comparison and assessment of RNA-SEQ procedures for gene expression quantitative analysis. *Scientific Reports*, 10(1), 1–15. <https://doi.org/10.1038/s41598-020-76881-x>
19. Corley, S. M., Troy, N. M., Bosco, A., & Wilkins, M. R. (2019). QuantSeq. 3' Sequencing combined with Salmon provides a fast, reliable approach for high throughput RNA expression analysis. *Scientific Reports*, 9(1), 1–15. <https://doi.org/10.1038/s41598-019-55434-x>
20. Costela-ruiz, V. J., Illescas-montes, R., Puerta-puerta, J. M., Ruiz, C., & Melguizo-rodríguez, L. (2020). *Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information . January.*
21. COVID-19: Clinical features - UpToDate. (n.d.). Retrieved October 21, 2021, from <https://www.uptodate.com/contents/covid-19-clinical-features>
22. Cseke, L. J., Wu, W., & Kaufman, P. B. (2003). DNA sequencing and analysis. *Handbook of Molecular and Cellular Methods in Biology and Medicine*, Second Edition, 2015(11), 237–270. <https://doi.org/10.1201/b11351-11>

23. Devrimci-Ozguven, H., Kundakci, N., Kumbasar, H., & Boyvat, A. (2000). The depression, anxiety, life satisfaction and affective expression levels in psoriasis patients. *Journal of the European Academy of Dermatology and Venereology*, 14(4), 267–271. <https://doi.org/10.1046/j.1468-3083.2000.00085.x>
24. *Differential gene expression (DGE) analysis | Training-modules*. (n.d.). Retrieved October 21, 2021, from https://hbctraining.github.io/Training-modules/planning_successful_rnaseq/lessons/sample_level_QC.html
25. Dobin, A. (2019). *Manual_STAR*. 1–50. <https://github.com/alexdobin/STAR/blob/master/doc/STARmanual.pdf>
26. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-SEQ aligner. *Bioinformatics*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
27. Fortune, D. G., Richards, H. L., Kirby, B., Bowcock, S., Main, C. J., & Griffiths, C. E. M. (2002). A cognitive-behavioural symptom management programme as an adjunct in psoriasis therapy. *British Journal of Dermatology*, 146(3), 458–465. <https://doi.org/10.1046/j.1365-2133.2002.04622.x>
28. *Gene Ontology Resource*. (n.d.). Retrieved October 28, 2021, from <http://geneontology.org/>
29. Gombor, S., Chang, M., Hogan, C. A., Zehnder, J., Boyd, S., Pinsky, B. A., & Shah, N. H. (2020). Persistent detection of SARS-CoV-2 RNA in patients and healthcare workers with COVID-19. *Journal of Clinical Virology*, 129(May), 104477. <https://doi.org/10.1016/j.jcv.2020.104477>
30. Gu, H., & Yuan, G. (2020). Identification of key genes in SARS-CoV-2 patients on bioinformatics analysis. *BioRxiv*, 2020.08.09.243444. <https://doi.org/10.1101/2020.08.09.243444>
31. Head, S. R., Kiyomi Komori, H., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: Overviews and

- challenges. *BioTechniques*, 56(2), 61–77. <https://doi.org/10.2144/000114133>
32. Heining, U. (2020). Severe acute respiratory syndrome coronavirus 2 vaccines: Setting expectations appropriately. *Pediatric Infectious Disease Journal*, January, E123–E124. <https://doi.org/10.1097/INF.0000000000002741>
33. Henderson, L. A., Canna, S. W., Schulert, G. S., Volpi, S., Lee, P. Y., Kernan, K. F., Caricchio, R., Mahmud, S., Hazen, M. M., Halyabar, O., Hoyt, K. J., Han, J., Grom, A. A., Gattorno, M., Ravelli, A., De Benedetti, F., Behrens, E. M., Cron, R. Q., & Nigrovic, P. A. (2020). On the Alert for Cytokine Storm: Immunopathology in COVID-19. *Arthritis and Rheumatology*, 72(7), 1059–1063. <https://doi.org/10.1002/art.41285>
34. Huang, A. T., Garcia-Carreras, B., Hitchings, M. D. T., Yang, B., Katzelnick, L. C., Rattigan, S. M., Borgert, B. A., Moreno, C. A., Solomon, B. D., Trimmer-Smith, L., Etienne, V., Rodriguez-Barraquer, I., Lessler, J., Salje, H., Burke, D. S., Wesolowski, A., & Cummings, D. A. T. (2020). A systematic review of antibody mediated immunity to coronaviruses: kinetics, correlates of protection, and association with severity. *Nature Communications*, 11(1), 1–16. <https://doi.org/10.1038/s41467-020-18450-4>
35. Islam, A. B. M. M. K., Khan, M. A. A. K., Ahmed, R., Hossain, M. S., Kabir, S. M. T., Islam, M. S., & Siddiki, A. M. A. M. Z. (2021). Transcriptome of nasopharyngeal samples from COVID-19 patients and a comparative analysis with other SARS-CoV-2 infection models reveal disparate host responses against SARS-CoV-2. *Journal of Translational Medicine*, 19(1), 1–25. <https://doi.org/10.1186/s12967-020-02695-0>
36. Jain, R., Ramaswamy, S., Harilal, D., Uddin, M., Loney, T., Nowotny, N., Alsuwaidi, H., Varghese, R., Deesi, Z., Alkhajeh, A., Khansaheb, H., Alsheikh-Ali, A., & Abou Tayoun, A. (2021). Host transcriptomic profiling of COVID-19 patients with mild, moderate, and severe clinical

- outcomes. *Computational and Structural Biotechnology Journal*, 19, 153–160. <https://doi.org/10.1016/j.csbj.2020.12.016>
37. Jean-F. Tomb, White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., ... Venter, J. C. (1997). Enhanced Reader.pdf. In *Nature* (Vol. 388, pp. 539–547).
38. Jiao, X., Sherman, B. T., Huang, D. W., Stephens, R., Baseler, M. W., Lane, H. C., & Lempicki, R. A. (2012). DAVID-WS: A stateful web service to facilitate gene/protein list analysis. *Bioinformatics*, 28(13), 1805–1806. <https://doi.org/10.1093/bioinformatics/bts251>
39. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065). <https://doi.org/10.1098/RSTA.2015.0202>
40. Jordan, S. C. (2021). Innate and adaptive immune responses to SARS-CoV-2 in humans: relevance to acquired immunity and vaccine responses. *Clinical and Experimental Immunology*, 204(3), 310. <https://doi.org/10.1111/CEI.13582>
41. Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1), 27–30. <https://doi.org/10.1093/NAR/28.1.27>
42. Kelleni, M. (2021). *NSAIDs Immunomodulation in COVID-19 Might Inhibit SARS CoV-2 ORF Proteins Induced Caspase Activation , Necroptosis and Endoplasmic Reticulum Stress . 522098.*
43. Kevadiya, B. D., Machhi, J., Herskovitz, J., Oleynikov, M. D., Blomberg, W. R., Bajwa, N., Soni, D., Das, S., Hasan, M., Patel, M., Senan, A. M., Gorantla, S., McMillan, J. E., Edagwa, B., Eisenberg, R., Gurumurthy, C. B., Reid, S. P. M., Punyadeera, C., Chang, L., & Gendelman, H. E. (2021). Diagnostics for SARS-CoV-2 infections. *Nature Materials*, 20(5), 593–605. <https://doi.org/10.1038/s41563-020-00906-z>

44. Kim, J. S., Lee, J. Y., Yang, J. W., Lee, K. H., Effenberger, M., Szpirt, W., Kronbichler, A., & Shin, J. II. (2020). Immunopathogenesis and treatment of cytokine storm in COVID-19. *Theranostics*, 11(1), 316–329. <https://doi.org/10.7150/THNO.49713>
45. Khanmohammadi, S., & Rezaei, N. (2021). Role of Toll-like receptors in the pathogenesis of COVID-19. *Journal of Medical Virology*, 93(5), 2735–2739. <https://doi.org/10.1002/jmv.26826>
46. Knoll, R., Schultze, J. L., & Schulte-Schrepping, J. (2021). Monocytes and Macrophages in COVID-19. *Frontiers in Immunology*, 12(July), 1–12. <https://doi.org/10.3389/fimmu.2021.720109>
47. Li, G., Wang, J., He, X., Zhang, L., Ran, Q., Xiong, A., Wu, D., Hu, L., Song, Q., & Zhu, D. (2020). An integrative analysis identifying transcriptional features and key genes involved in COVID-19. *Epigenomics*, 12(22), 1969–1981. <https://doi.org/10.2217/epi-2020-0168>
48. Li, W. V., & Li, J. J. (2018). Modeling and analysis of RNA-seq data: a review from a statistical perspective. *Quantitative Biology (Beijing, China)*, 6(3), 195. <https://doi.org/10.1007/S40484-018-0144-7>
49. Liang, H., & Zeng, E. (2016). RNA-SEQ Experiment and Data Analysis. *Methods in Molecular Biology*, 1366, 99–114. https://doi.org/10.1007/978-1-4939-3127-9_9
50. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-SEQ data with DESeq2. *Genome Biology*, 15(12), 1–21. <https://doi.org/10.1186/s13059-014-0550-8>
51. M, S., V, A., T, R., M, S., S, D., M, R., D, B., H, L., & ML, Y. (2014). Influence of RNA extraction methods and library selection schemes on RNA-SEQ data. *BMC Genomics*, 15(1). <https://doi.org/10.1186/1471-2164-15-675>
52. Ma, F., Fuqua, B. K., Hasin, Y., Yukhtman, C., Vulpe, C. D., Lusic, A. J., & Pellegrini, M. (2019). A comparison between whole transcript and 3' RNA sequencing methods using Kapa and Lexogen library preparation methods. *BMC Genomics* 2019 20:1, 20(1), 1–12. <https://doi.org/10.1186/S12864-018-5393-3>

53. Machhi, J., Herskovitz, J., Senan, A. M., Dutta, D., Nath, B., Oleynikov, M. D., Blomberg, W. R., Meigs, D. D., Hasan, M., Patel, M., Kline, P., Chang, R. C. C., Chang, L., Gendelman, H. E., & Kevadiya, B. D. (2020). The Natural History, Pathobiology, and Clinical Manifestations of SARS-CoV-2 Infections. *Journal of Neuroimmune Pharmacology*, 15(3), 359–386. <https://doi.org/10.1007/s11481-020-09944-5>
54. Madewell, Z. J., Yang, Y., Jr, I. M. L., Halloran, M. E., & Dean, N. E. (2020). NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice. 1. *MedRxiv*, 6(165), 1–13.
55. Mandal, S., Barnett, J., Brill, S. E., Brown, J. S., Denny, E. K., Hare, S. S., Heightman, M., Hillman, T. E., Jacob, J., Jarvis, H. C., Lipman, M. C. I., Naidu, S. B., Nair, A., Porter, J. C., Tomlinson, G. S., Hurst, J. R., & Group, A. S. (2020). ‘Long-COVID’: a cross-sectional study of persisting symptoms, biomarker and imaging abnormalities following hospitalisation for COVID-19. *Thorax*, 76(4), 396–398. <https://doi.org/10.1136/THORAXJNL-2020-215818>
56. Melo, A. K. G., Milby, K. M., Caparroz, A. L. M. A., Pinto, A. C. P. N., Santos, R. R. P., Rocha, A. P., Ferreira, G. A., Souza, V. A., Valadares, L. D. A., Vieira, R. M. R. A., Pileggi, G. S., & Trevisani, V. F. M. (2021). Biomarkers of cytokine storm as red flags for severe and fatal COVID-19 cases: A living systematic review and meta-analysis. *PLOS ONE*, 16(6), e0253894. <https://doi.org/10.1371/JOURNAL.PONE.0253894>
57. Melo, C. V. B., Bhuiyan, M. A., Gatua, W. N., Kanyerezi, S., Uzairue, L., Abechi, P., Kumar, K., Rahmat, J., Giwa, A., Mwandira, G., Olamilekan, A. M., George, T. E., Adejinmi, O. J., Ibrinke, M. A., Rotimi, O. D., Abo-Elenein, D. A. M. A., Abubakar, R. A., Usman, M., Adewunmi, I., ... Carl, S. H. (2020). *Transcriptomic dysregulations associated with SARS-CoV-2 infection in human nasopharyngeal and peripheral blood mononuclear cells*. <https://doi.org/10.1101/2020.09.09.289850>

58. Michael, A., Ahlmann-eltze, C., Forbes, K., & Anders, S. (2021). *Package 'DESeq2'.*
59. Mostafa, H. H., Fissel, J. A., Fanelli, B., Bergman, Y., Gniazdowski, V., Dadlani, M., Carroll, K. C., Colwell, R. R., & Simner, P. J. (2020). Metagenomic next-generation sequencing of nasopharyngeal specimens collected from confirmed and suspect covid-19 patients. *MBio*, 11(6), 1–13. <https://doi.org/10.1128/mBio.01969-20>
60. Mueller, S. N., & Rouse, B. T. (2008). Immune responses to viruses. *Clinical Immunology, January*, 421–431. <https://doi.org/10.1016/B978-0-323-04404-2.10027-2>
61. Mun, D. G., Vanderboom, P. M., Madugundu, A. K., Garapati, K., Chavan, S., Peterson, J. A., Saraswat, M., & Pandey, A. (2021). DIA-Based Proteome Profiling of Nasopharyngeal Swabs from COVID-19 Patients. In *Journal of Proteome Research* (Vol. 20, Issue 8, pp. 4165–4175). <https://doi.org/10.1021/acs.jproteome.1c00506>
62. Nalbandian, A., Sehgal, K., Gupta, A., Madhavan, M. V., McGroder, C., Stevens, J. S., Cook, J. R., Nordvig, A. S., Shalev, D., Sehwat, T. S., Ahluwalia, N., Bikdeli, B., Dietz, D., Der-Nigoghossian, C., Liyanage-Don, N., Rosner, G. F., Bernstein, E. J., Mohan, S., Beckley, A. A., ... Wan, E. Y. (2021). Post-acute COVID-19 syndrome. *Nature Medicine* 2021 27:4, 27(4), 601–615. <https://doi.org/10.1038/s41591-021-01283-z>
63. Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., Khodadoust, M. S., Esfahani, M. S., Luca, B. A., Steiner, D., Diehn, M., & Alizadeh, A. A. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nature Biotechnology*, 37(7), 773–782. <https://doi.org/10.1038/s41587-019-0114-2>
64. Ng, D. L., Granados, A. C., Santos, Y. A., Servellita, V., Goldgof, G. M., Meydan, C., Sotomayor-Gonzalez, A., Levine, A. G., Balcerak, J., Han, L. M., Akagi, N., Truong, K., Neumann, N. M., Nguyen, D. N., Bapat, S. P., Cheng, J., Martin, C. S. S., Federman, S., Fook, J., ... Chiu, C. Y. (2021). A diagnostic host response biosignature for COVID-19 from

- RNA profiling of nasal swabs and blood. *Science Advances*, 7(6), 1–16. <https://doi.org/10.1126/sciadv.abe5984>
65. Nour, E. T., Tran, R., Afravi, A., Pei, X., & Davidian, A. (2021). *Identifying The “ Core ” Transcriptome of SARS-CoV-2 Infected Cells*. 1–22.
66. O’Connell, P., & Aldhamen, Y. A. (2020). Systemic innate and adaptive immune responses to SARS-CoV-2 as it relates to other coronaviruses. <https://doi.org/10.1080/21645515.2020.1802974>, 16(12), 2980–2991. <https://doi.org/10.1080/21645515.2020.1802974>
67. Or Caspi, Michael J. Smart, R. B. N. (2020). Adaptive immunity to SARS-CoV-2 and COVID-19 Alessandro. *Ann Oncol, January*, 19–21.
68. Pan, Y., Jiang, X., Yang, L., Chen, L., Zeng, X., Liu, G., Tang, Y., Qian, C., Wang, X., Cheng, F., Lin, J., Wang, X., & Li, Y. (2021). SARS-CoV-2-specific immune response in COVID-19 convalescent individuals. *Signal Transduction and Targeted Therapy*, 6(1). <https://doi.org/10.1038/s41392-021-00686-1>
69. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14(4), 417–419. <https://doi.org/10.1038/NMETH.4197>.
70. Picardi, A., & Abeni, D. (2001). Stressful life events and skin diseases: Disentangling evidence from myth. *Psychotherapy and Psychosomatics*, 70(3), 118–136. <https://doi.org/10.1159/000056237>
71. Qi, F., Xu, G., Liao, X., Wang, F., Yuan, J., Wang, H., Wang, X., Liu, L., Zhang, S., & Zhang, Z. (2021). ScRNA-SEQ revealed the kinetic of nasopharyngeal immune responses in asymptomatic COVID-19 carriers. *Cell Discovery*, 7(1), 0–3. <https://doi.org/10.1038/s41421-021-00294-x>
72. Ragab, D., Salah Eldin, H., Taeimah, M., Khattab, R., & Salem, R. (2020). The COVID-19 Cytokine Storm; What We Know So Far. *Frontiers in Immunology*, 11(June), 1–4. <https://doi.org/10.3389/fimmu.2020.01446>

73. Ren, X., Wen, W., Fan, X., Hou, W., Su, B., Cai, P., Li, J., Liu, Y., Tang, F., Zhang, F., Yang, Y., He, J., Ma, W., He, J., Wang, P., Cao, Q., Chen, F., Chen, Y., Cheng, X., ... Zhang, Z. (2021). COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell*, *184*(7), 1895-1913.e19. <https://doi.org/10.1016/j.cell.2021.01.053>
74. Ricci, D., Etna, M. P., Rizzo, F., Sandini, S., Severa, M., & Coccia, E. M. (2021). Innate Immune Response to SARS-CoV-2 Infection: From Cells to Soluble Mediators. *International Journal of Molecular Sciences*, *22*(13). <https://doi.org/10.3390/IJMS22137017>
75. Risso, D., Ngai, J., Speed, T. P., & Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, *32*(9), 896–902. <https://doi.org/10.1038/nbt.2931>
76. Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
77. Rosa, B. A., Ahmed, M., Singh, D. K., Choreño-Parra, J. A., Cole, J., Jiménez-Álvarez, L. A., Rodríguez-Reyna, T. S., Singh, B., Gonzalez, O., Carrion, R., Schlesinger, L. S., Martin, J., Zúñiga, J., Mitreva, M., Kaushal, D., & Khader, S. A. (2021). IFN signaling and neutrophil degranulation transcriptional signatures are induced during SARS-CoV-2 infection. *Communications Biology*, *4*(1), 1–14. <https://doi.org/10.1038/s42003-021-01829-4>
78. Rusinova, I., Forster, S., Yu, S., Kannan, A., Masse, M., Cumming, H., Chapman, R., & Hertzog, P. J. (2013). INTERFEROME v2.0: An updated database of annotated interferon-regulated genes. *Nucleic Acids Research*, *41*(D1), 1040–1046. <https://doi.org/10.1093/nar/gks1215>
79. Saheb Sharif-Askari, N., Saheb Sharif-Askari, F., Mdkhana, B., Al Heialy, S., Alsafar, H. S., Hamoudi, R., Hamid, Q., & Halwani, R. (2021). Enhanced expression of immune checkpoint receptors during

- SARS-CoV-2 viral infection. *Molecular Therapy - Methods and Clinical Development*, 20(March), 109–121. <https://doi.org/10.1016/j.omtm.2020.11.002>
80. Schaarschmidt, S., Fischer, A., Zuther, E., & Hinch, D. K. (2020). Evaluation of seven different RNA-SEQ alignment tools based on experimental data from the model plant arabidopsis thaliana. *International Journal of Molecular Sciences*, 21(5). <https://doi.org/10.3390/ijms21051720>
81. Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., & Ragg, T. (2006). The RIN: An RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology*, 7, 1–14. <https://doi.org/10.1186/1471-2199-7-3>
82. Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G. G., Owen-Hughes, T., Blaxter, M., & Barton, G. J. (2016). How many biological replicates are needed in an RNA-SEQ experiment and which differential expression tool should you use? *Rna*, 22(6), 839–851. <https://doi.org/10.1261/rna.053959.115>
83. Schwartz, J., Evers, A. W. M., Bundy, C., & Kimball, A. B. (2016). Getting under the Skin: Report from the International Psoriasis Council Workshop on the Role of Stress in Psoriasis. *Frontiers in Psychology*, 7(February), 1–4. <https://doi.org/10.3389/fpsyg.2016.00087>
84. Singh, K., Chen, Y. C., Hassanzadeh, S., Han, K., Judy, J. T., Seifuddin, F., Tunc, I., Sack, M. N., & Pirooznia, M. (2021). Network Analysis and Transcriptome Profiling Identify Autophagic and Mitochondrial Dysfunctions in SARS-CoV-2 Infection. *Frontiers in Genetics*, 12(March), 1–19. <https://doi.org/10.3389/fgene.2021.599261>
85. Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nature Reviews Genetics*, 20(11), 631–656. <https://doi.org/10.1038/s41576-019-0150-2>

86. Su, C. M., Wang, L., & Yoo, D. (2021). Activation of NF-κB and induction of proinflammatory cytokine expressions mediated by ORF7a protein of SARS-CoV-2. *Scientific Reports*, 11(1), 1–12. <https://doi.org/10.1038/s41598-021-92941-2>
87. Sungnak, W., Huang, N., Bécavin, C., Berg, M., Queen, R., Litvinukova, M., Talavera-López, C., Maatz, H., Reichart, D., Sampaziotis, F., Worlock, K. B., Yoshida, M., Barnes, J. L., Banovich, N. E., Barbry, P., Brazma, A., Collin, J., Desai, T. J., Duong, T. E., ... Figueiredo, F. (2020). SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nature Medicine*, 26(5), 681–687. <https://doi.org/10.1038/s41591-020-0868-6>
88. Trisha Greenhalgh, Matthew Knight, Christine A'Court, Maria Buxton, L. H. (2020). *Management of post-acute covid-19 in primary care – Enhanced Reader.pdf*.
89. Vafaeinezhad, A., Atashzar, M. R., & Baharlou, R. (2021). The Immune Responses against Coronavirus Infections: Friend or Foe? *International Archives of Allergy and Immunology*, 182(9), 863–876. <https://doi.org/10.1159/000516038>
90. van Doremalen, N., Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., Tamin, A., Harcourt, J. L., Thornburg, N. J., Gerber, S. I., Lloyd-Smith, J. O., de Wit, E., & Munster, V. J. (2020). Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. In *New England Journal of Medicine* (Vol. 382, Issue 16, pp. 1564–1567). <https://doi.org/10.1056/nejmc2004973>
91. Vaziri, T., Michalsky, S., Premeaux, T. A., Pang, A. P. S., Bukhari, Z., Rindone, N., Loizou, D., Ndhlovu, L. C., & Gupta, R. (2020). 5,#,* 1 2.
92. Vilas, C., Melo, B. De, Bhuiyan, M. A., Gatua, W. N., Kanyerezi, S., Uzairue, L., Abechi, P., Kumar, K., Rahmat, J., Giwa, A., Mwandira, G., Olamilekan, A. M., George, T. E., Adejinmi, O. J., Ibrinke, M. A., Rotimi, O. D., Aly, D., Abo-elenein, M. A., Abubakar, R. A., ... Carl, S. H. (2020). *Nasopharyngeal and Peripheral Blood Mononuclear Cells*. doi: 10.1080/15384047.2019.1670521.

93. Warren, H. S. (2005). Toll-like receptors. *Critical Care Medicine*, 33(12 SUPPL.). <https://doi.org/10.1097/01.CCM.0000185504.39347.5D>
94. Williams, A. G., Thomas, S., Wyman, S. K., & Holloway, A. K. (2014). RNA-seq Data: Challenges in and Recommendations for Experimental Design and Analysis. In *Current protocols in human genetics* (Vol. 83, pp. 11.13.1-11.13.20). <https://doi.org/10.1002/0471142905.hg1113s83>
95. Y, H., Y, Z., & L, M. (2021). Meta-analysis of laboratory results in patients with severe coronavirus disease 2019. *Experimental and Therapeutic Medicine*, 21(5). <https://doi.org/10.3892/ETM.2021.9877>
96. Yalamanchili, H. K., Wan, Y. W., & Liu, Z. (2017). Data analysis pipeline for RNA-seq experiments: From differential expression to cryptic splicing. *Current Protocols in Bioinformatics*, 2017, 11.15.1-11.15.21. <https://doi.org/10.1002/cpbi.33>
97. Yang, L., Xie, X., Tu, Z., Fu, J., Xu, D., & Zhou, Y. (2021). The signal pathways and treatment of cytokine storm in COVID-19. *Signal Transduction and Targeted Therapy*, 6(1), 1–20. <https://doi.org/10.1038/s41392-021-00679-0>
98. Zhang, C. D., Yang, Y., Chen, H. H., Zhang, T., Wang, Q., Liang, Y., Zhang, L., & Zhou, Y. (2018). RTPDB: A database providing associations between genetic variation or expression and cancer prognosis with radiotherapy-based treatment. *Database*, 2018(2018), 1–6. <https://doi.org/10.1093/database/bay118>
99. Zhang, Q., Meng, Y., Wang, K., Zhang, X., Chen, W., Sheng, J., Qiu, Y., Diao, H., & Li, L. (2021). Inflammation and Antiviral Immune Response Associated With Severe Progression of COVID-19. *Frontiers in Immunology*, 0, 135. <https://doi.org/10.3389/FIMMU.2021.631226>
100. Zohar, T., & Alter, G. (2020). Dissecting antibody-mediated protection against SARS-CoV-2. *Nature Reviews Immunology*, 20(7), 392–394. <https://doi.org/10.1038/s41577-020-0359-5>

6.2 ΕΛΛΗΝΟΦΩΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Αποθετήριο Κάλλιπος: Υπολογιστική βιολογία. (n.d.). Retrieved October 21, 2021, from <https://repository.kallipos.gr/handle/11419/1577>