NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF SCIENCES

DEPARTMENT OF CHEMISTRY

DOCTORAL THESIS

Development of methodologies for the identification of biomarkers in clinical/ biological samples by high resolution mass spectrometry techniques

BARLA V. IOANNA
MSc. CHEMIST

ATHENS

DECEMBER 2023

DOCTORAL THESIS

Development of methodologies for the identification of biomarkers in clinical/ biological samples by high resolution mass spectrometry techniques

IOANNA BARLA

REGISTRATION NUMBER: 2028

SUPERVISING PROFESSOR: Dr. Evagelos Gikas, Professor

THREE-MEMBER CONSULTATIVE COMMITTEE:

Dr. Evagelos Gikas, Professor

Dr. Nikolaos S. Thomaidis, Professor

Dr. Ioannis Dotsikas, Associate Professor

SEVEN-MEMBER EXAMINATION COMMITTEE

Dr. Evagelos Gikas, Professor

Dr. Nikolaos S. Thomaidis, Professor

Dr. Ioannis Dotsikas, Associate Professor

Dr. Antony Tsarbopoulos, Professor

Dr. George Panayotou, Researcher A’

Dr. Eleni Gikas, Assistant Professor

Dr. Guro Fanneløb Giskeødegård, Associate Professor

DEFENDING DATE 04/12/2023

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Ανάπτυξη μεθοδολογιών για την ταυτοποίηση βιοδεικτών σε κλινικά/ βιολογικά δείγματα με τη χρήση Φασματομετρίας Μάζας Υψηλής Διακριτικής Ικανότητας

ΙΩΑΝΝΑ ΜΠΑΡΛΑ

Α.Μ.: 2028

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: Δρ. Ευάγγελος Γκίκας, Καθηγητής

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ:

Δρ. Ευάγγελος Γκίκας, Καθηγητής

Δρ. Νικόλαος Σ. Θωμαΐδης, Καθηγητής

Δρ. Ιωάννης Ντότσικας, Αναπληρωτής Καθηγητής

ΕΠΤΑΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

Δρ. Ευάγγελος Γκίκας, Καθηγητής

Δρ. Νικόλαος Σ. Θωμαΐδης, Καθηγητής

Δρ. Ιωάννης Ντότσικας, Αναπληρωτής Καθηγητής

Δρ. Αντώνιος Τσαρμπόπουλος, Καθηγητής

Δρ. Γεώργιος Παναγιώτου, Ερευνητής Α’

Δρ. Ελένη Γκίκα, Επίκουρη Καθηγήτρια

Dr. Guro Fanneløb Giskeødegård, Αναπληρώτρια Καθηγήτρια

ΗΜΕΡΟΜΗΝΙΑ ΕΞΕΤΑΣΗΣ 04/12/2023

Περίληψη

Σημαντικός στόχος της σύγχρονης τοξικολογίας είναι η εκτίμηση της αποτελεσματικότητας και των παρενεργειών (όπως η τοξικότητα) των φαρμακευτικών ενώσεων, πριν από το στάδιο των κλινικών δοκιμών. Για αυτό εστιάζει στην ανάπτυξη γρήγορων, πολύμορφων και φθηνών μεθόδων, που θα επιτρέπουν την πρόβλεψη της τοξικότητας κατά την διάρκεια των *in-vivo* δοκιμασιών. Οι νέες αναλυτικές τεχνικές και τα αναδυόμενα υπολογιστικά εργαλεία διευκολύνουν την παραπάνω προσπάθεια. Οι μεταβολομικές μελέτες που βασίζονται σε φασματομετρία μάζας υψηλής διακριτικής ικανότητας (HRMS) παρουσιάζουν ιδιαίτερο ενδιαφέρον για τις εφαρμογές της σύγχρονης τοξικολογίας. Οι μεταβολίτες είναι ενώσεις που έχουν καθολικές δομές και επιτρέπουν την μετάφραση της βιολογικής απόκρισης από το επίπεδο του πειραματόζωου στο επίπεδο του ανθρώπινου οργανισμού. Επίσης, το μεταβόλωμα περιγράφει με ακρίβεια έναν φαινότυπο ενδιαφέροντος και επομένως η μεταβολομική μελέτη επιτρέπει την βαθύτερη κατανόηση της βιοχημείας των εκφράσεων τοξικότητας, καθώς και την δημιουργία προβλεπτικών μοντέλων τοξικότητας. Βάσει των παραπάνω, η παρούσα διδακτορική διατριβή εστιάζει στην ανάπτυξη μεθοδολογιών μεταβολομικής ανάλυσης για την εύρεση βιοδεικτών που σχετίζονται με την έκφραση τοξικότητας.

Το πρώτο μέρος της παρούσας διατριβής (Κεφάλαια 4 και 5) παρουσιάζει την διερεύνηση της νεφροτοξικότητας που επάγεται από την χορήγηση του αντικαρκινικού παράγοντα καρφιλζομίμπης (Cfz). Κατά το *in-vivo* πείραμα, σε 6 ποντίκια χορηγήθηκαν 8 mg/kg/μέρα Cfz (Cfz-group) και σε 6 ποντίκια χορηγήθηκε φυσιολογικός ορός (Control-group). Το πλάσμα, οι νεφροί και τα ούρα των ποντικών αναλύθηκαν με UPLC-HRMS-DIA μεθοδολογία. Για την ταξινόμηση των ομάδων και την επιλογή μεταβλητών χρησιμοποιήθηκε συνδυασμός πολυπαραμετρικών και μονοπαραμετρικών μεθοδολογιών. Για την ταυτοποίηση των μεταβολητών χρησιμοποιήθηκαν μεθοδολογίες μη στοχευμένης σάρωσης και σάρωσης ύποπτων ενώσεων. Επίσης, αναπτύχθηκαν post-hoc μεθοδολογίες για την διερεύνηση συσχετίσεων και αλληλεπιδράσεων ανάμεσα στους μεταβολίτες και τους διαφορετικούς τύπους βιοδειγμάτων. Η μελέτη απέδειξε την σοβαρή επίπτωση του φαρμάκου στους

νεφρούς, μέσω της διατάραξης του νεφρικού μεταβολισμού που οδηγεί σε νεφρική δυσλειτουργία. Επιπλέον, ταυτοποιήθηκαν μεταβολίτες που σχετίζονται με την ουρεμία, το οξειδωτικό στρες, την φλεγμονή και την νεφρική ανεπάρκεια.

Το δεύτερο μέρος της μελέτης (Κεφάλαιο 6) εστίασε στον προσδιορισμό πρώιμων αλλαγών στο μεταβολικό προφίλ παιδιών με καρκίνο, οι οποίες θα μπορούσαν να εξηγούν/ προβλέπουν τον κίνδυνο εμφάνισης καρδιοτοξικότητας, όταν οι ασθενείς υποβληθούν σε θεραπεία. Σε συνεργασία με την Ογκολογική – Αιματολογική Κλινική του Νοσοκομείου Παίδων «Αγία Σοφία» πραγματοποιήθηκαν δειγματοληψίες αίματος από ασθενείς, πριν την υποβολή τους σε χημειοθεραπεία. Οι ασθενείς ακολούθησαν το κατάλληλο θεραπευτικό πρωτόκολλο και κάποιοι (26) εμφάνισαν καρδιοτοξικότητα. Δείγματα πλάσματος αναλύθηκαν με UPLC-HRMS-DIA μεταβολομική ανάλυση και, η γνώση εμφάνισης τοξικότητας χρησιμοποιήθηκε για την ταξινόμηση των ασθενών σε Ομάδα Ρίσκου (CT-Risk) και ομάδα ελέγχου (No-Risk). Η μελέτη εστίασε στην διαδικασία επιλογής μεταβλητών συνδυάζοντας 4 συμπληρωματικές στατιστικές δοκιμασίες (KODAMA, OPLS-DA, BORUTA, t-test). Η βασική ιδέα ήταν πως μεταβλητές με καλή απόδοση σε παραπάνω από 3 δοκιμασίες έχουν αυξημένη πιθανότητα να σχετίζονται με το ρίσκο εμφάνισης καρδιοτοξικότητας, σύμφωνα με την θεωρεία του Bayes. Επιπλέον, η μελέτη έθεσε συγκεκριμένα κριτήρια εμπιστοσύνης για την ταυτοποίηση μεταβολιτών που προκύπτουν από DIA μεθοδολογίες λήψης φασμάτων. Η μελέτη ανέπτυξε μοντέλα ταυτοποίησης των CT-Risk ασθενών, με επιβεβαιωμένη προβλεπτική ικανότητα. Επιπλέον, οι ταυτοποιημένοι μεταβολίτες έδειξαν πρώιμη μεταβολή σε μεταβολικά μονοπάτια που συνδέονται με το γενικό μονοπάτι του καρδιακού μεταβολισμού για παραγωγή ενέργειας. Αυτές οι παρατηρήσεις υποδεικνύουν πως η χημειοθεραπεία επιδεινώνει ήδη υπάρχουσες ανωμαλίες της καρδιακής λειτουργίας και οδηγεί στην εκδήλωση καρδιοτοξικότητας.

Τελευταίο μέρος της παρούσας διατριβής (Κεφάλαιο 7) αποτέλεσε η διερεύνηση της χορήγησης ανθρώπινης δόσης κολιστίνης σε ποντίκια. Στόχος ήταν η κατανόηση των βιοχημικών αιτιών που οδηγούν σε τοξικότητα λόγο του συγκεκριμένου φαρμάκου. Η κολιστίνη (CMS) είναι αντιβιοτικό τελευταίας επιλογής, και συνδέεται με σοβαρή νεύρο- και νέφροτοξικότητα. Θέλοντας να δούμε την επίπτωση του φαρμάκου στον οργανισμό σε συνθήκες που δεν

υπάρχουν κλινικές ενδείξεις τοξικότητας, πραγματοποιήθηκε ένα *in-vivo* πείραμα: σε μία ομάδα ποντικών χορηγήθηκε 1 mg/kg/ ημέρα (LD-Group), σε μία ομάδα ποντικών χορηγήθηκαν 1.5 mg/kg/ ημέρα (HD-Group), και σε μία ομάδα ποντικών χορηγήθηκε φυσιολογικός ορός (Control). Το πλάσμα, οι νεφροί και το ήπαρ, χρησιμοποιήθηκαν για μεταβολομική ανάλυση με τη χρήση UPLC-HRMS-DIA μεθοδολογίας. Τα αποτελέσματα ταυτοποίησαν μεταβολίτες που ανταποκρίνονται στην αύξηση της δόσης. Επίσης εντοπίστηκε απορρύθμιση του μεταβολισμού της ντοπαμίνης και του μεταβολισμού των πουρινών στους νεφρούς. Ιδιαίτερη σημασία είχε η αύξηση της ξανθίνης στο επίπεδο των νεφρών, που επάγει τη δράση του ενζύμου ακετυλοχολινεστεράση, οδηγώντας σε ταχεία αποικοδόμηση της ακετυλοχολίνης, διαδικασία που θα μπορούσε να αποτελέσει συνδετικό κρίκο ανάμεσα στην νευροτοξικότητα και την νεφροτοξικότητα λόγω της χορήγησης κολιστίνης.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Χημική Ανάλυση

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: μεταβολομική, φασματομετρία μάζας υψηλής διακριτικής ικανότητας (HRMS), βιοδείκτες, τοξικολογία, χημειομετρία, επιλογή μεταβλητών, κολιστίνη, καρφιλζομίμπη, παιδιά με νεοπλασίες, νεφροτοξικότητα, καρδιοτοξικότητα, μελέτη ρίσκου εμφάνισης τοξικότητα

Abstract

The major goal of 21st century toxicology is to assess the efficiency/ toxicity of drugs before the step of clinical trials, and therefore focuses on the development of quick, versatile, and cheap methods that will enable the toxicity prediction during the in-vivo testing. The novel analytical methodologies and the emerging computational resources facilitate this effort. HRMS-based metabolomics present a particular interest in the terms of modern toxicology. The metabolites have universal structures and therefore, they facilitate the animal-to-human translation. Also, the metabolome is a consistent descriptor of the phenotype, so the metabolomics studies permit the deep comprehension of the biochemistry causes of toxicity and allow the creation of prediction models as well. Under this notion the current doctoral thesis focuses on the development of thorough metabolomics methodologies in the terms of the detection of toxicity related biomarkers.

The first part of this study (Chapters 4&5) presents the investigation of the nephrotoxicity caused by the administration of the antineoplastic drug Carfilzomib (Cfz). An *in-vivo* experiment was performed, using six mice that were treated with 8mg/kg Cfz (Cfz-group) and six mice that received normal saline (Control-group). The plasma, the kidneys and the urine were used for UPLC-HRMS-DIA-based metabolomics analysis. Multivariate and univariate chemometrics were implemented for the classification and the variable selection. In addition, library-free and library-based methods were used for metabolites identification. The study developed post-hoc analysis methods to investigate potential correlations between the metabolites and to highlight interaction between the Cfz-toxicity (phenotype) and the circulatory/ urinary system (biosample). The study showed that the Cfz has severe impact on the kidneys and disrupts the renal metabolism, inducing kidney injury. Several of the identified metabolites were related to uremic condition, oxidative stress, inflammation, and kidney impairment.

The second part of this study (Chapter 6) focused on the detection of early alterations of the metabolomic profiles of oncology children's patients, that could explain/ predict the future expression of cardiotoxicity (CT), when they

would undergo chemotherapy. In collaboration with the Department of Oncology and Hematology of the Children's Hospital "Agia Sofia", bloods samples of patients were collected before their submission to chemotherapy. The children followed the appropriate therapeutic protocol and some expressed acute cardiotoxicity. A UPLC-HRMS-DIA metabolomics analysis was performed in the blood samples, and the *a posteriori* knowledge of the CT expression was used to group the children in CT-Risk and No-Risk. The investigation focused on the variable selection procedure, combining 4 complementary statistical methods (KODAMA, OPLS-DA, BORUTA, t-test). The motivating idea was that if a variable exhibits good performance in more than three of these tests, then it has increased probability to be CT-related (Bayesian Probabilistic Theory). Moreover, the study set specific criteria for the confidence of DIA-identification. Finally, the study achieved the classification of CT-risk patients with models that showed acceptable figures of merit. Moreover, identified metabolites showed early alterations in the metabolic pathways that participate in the cardiac energy metabolism and suggested that the chemotherapy triggers preexisting cardiac function abnormalities, leading to acute CT events.

The last part of the study (Chapter 7) investigated the impact of the human doses of colistin (CMS) in mice, aiming to shed light on the biochemical reasons of CMS toxicity. CMS is a last resort antibiotic related with severe neuro and nephrotoxicity. Aiming to shed light on the impact of CMS in non-toxic conditions, in the current study, two doses of CMS, Low (1 mg/kg) and High (1.5 mg/kg) versus a control (normal saline), were administered to mice. Samples of plasma, kidney, and liver were analyzed with a UPLC-HRMS-DIA-based metabolomics workflow. The data were submitted to PLS-DA, PLS-R and ROC analysis. The results pointed out six dose-responding metabolites, renal dopamine dysregulation, and extended perturbations in renal purine metabolism. An intriguing finding was the increased formation of renal xanthine, which is an AChE activator, leading to rapid degradation of achetylcholine, suggesting an association of nephrotoxicity and neurotoxicity.

SUBJECT AREA: Chemical Analysis

KEYWORDS: metabolomics, High resolution mass spectrometry (HRMS), biomarkers, toxicology, chemometrics, variable selection, colistin, carfilzomib, children with malignancies, nephrotoxicity, cardiotoxicity, toxicity risk assessment

ACKNOWLEDGMENTS

I am deeply indebted to my supervisor, Prof. Evagelos Gikas, for all his support, mentoring, and friendship throughout my PhD studies and for the opportunity to work with him and extend my scientific knowledge.

I would like to acknowledge the other members of the consultative committee, Prof. Ioannis Dotsikas and Prof. Nikolaos Thomaidis, for their guidance and insightful comments throughout this thesis. Furthermore, I would like to express my gratitude to Prof. Nikolaos Thomaidis for his trust, his support, and his mentoring all these years and for the opportunity to be a member of the TrAMS group.

I would like to thank the seven-member examination committee for their acceptance to participate in my PhD defense and the feedback they provided to improve the quality of my study.

I would like to express my deepest gratitude to Prof. Ioanna Andreadou and Dr. Sofia Polychronopoulou for their collaboration and their trust.

Of course, I would like to thank all my associates, my scientific friends, and the members of the TrAMS group for their collaboration, their encouragement, and their friendship.

I express my deepest gratitude to FLOGA, the Parents' Association of Children with Cancer, for the financial support.

Last but not least, I would like to thank my family and my friends for believing in me and for their support and encouragement all these years.

To my beloved people...

Table of Contents

List of Tables

List of Figures

s

1. Chapter 1

Drug safety and toxicology: State-of-the-art and modern approaches

1.1. Introduction

The assessment of the benefits and risks of a new drug and the efficient estimation of the adverse drug events, ADEs, currently represent a demanding challenge for the drug safety surveillance. According to the European Medicines Agency, EMA, the United States Food and Drug Administration, FDA, the China Food and Drug Administration, CFDA, and the Drugs Controller General of India, DCGI, the drug development process includes 5 main steps:

Step 1: Discovery and Development

Step 2: Preclinical Research

Step 3: Clinical Research

Step 4: FDA Review

Step 5: FDA Post-Market monitoring

The 1st step emphasizes in the discovery of new compounds that could be appropriate for medical treatment and, conducting experiments to gather information regarding compound's properties as the pharmacokinetics, the mechanism of action, the effectiveness, the optimal dosage regimen, the potential toxicity, as well as interactions with other drugs. The 2nd step includes in vitro and in vivo testing of the drug. The results are reviewed in order to decide if the new chemical entity can be further tested to humans. The 3rd step has 3 phases, described in Table 1. In this step the drug is administered in healthy volunteers, or patients of bearing a specific health condition, to assess their efficacy, ensuring their fit of purpose, and also to investigate potential toxic responses in humans. At the 4th Step, all the resulting data are reviewed by the responsible agencies, i.e., FDA, EMA, etc., and the final decision of approval or disapproval is made. After the release of a drug, the agencies continue to review reports on issues with prescription and over-the-counter drugs and could decide to add cautions concerning the dosage or use information, as well as other measures concerning more serious issues, (5th Step).

Table 1 The three phases of clinical trials

	Cohort	Duration	Purpose	Success rate
Phase I	20-100 healthy volunteers, or people with the disease/ condition	Several months	safety & the dosage	70%
Phase II	100-1000 hundreds of people with the disease/ condition	Months – 2 years	Efficacy & adverse effects monitoring	33%
Phase III	300-3000 of people with the disease/ condition	1-4 years	safety and efficacy	25-30%

The timeline of a drug development procedure lasts 12-15 years (https://phrma.org/policy-issues/Research-and-Development-Policy-Framework) until the final releasing approval. The cost to bring a new drug, from the new molecular entity, NME, in the clinical use rates from $160 million to $4.5 billion [1]. It is worth mentioning that less than 1% of the new chemical entities i.e., the compounds with potential therapeutic action discovered in the 1st step, are finally released as drugs. However, the rejection of a chemical compound in the first 2 steps of the drug development is not so resource-damaging, as the most time-consuming and expensive step are the clinical trials. Therefore, the major goal is to reject a compound as fast as possible in order to save time and resources.

Nevertheless, ~70% of the drugs that have succeed in the preclinical testing do not pass the clinical trials, showing unspecified toxicities or low efficacy (https://www.fda.gov/patients/drug-development-process/step-1-discovery-and-development). This is essential for both financial and ethical reasons, as the unspecified toxicities pose a high risk to the volunteers' health.

Even after the final approval, the drugs cannot be considered as safe, since the total picture of their safety is shaped after several months or years of use, when

they are tried to a large and random cohort of different phenotypes, demographics lifestyle, that may receive other drugs also etc. [2].

1.2. Toxicology in drug safety: Limitations & Prospectives

The rejection of drugs in the late phase of the clinical trials and the high uncertainty of their safety, create a major gap between the drug development and the increasing needs of effective and safe drugs. Thus, the observed low clinical trial success rates pose questions for the effectiveness of the so-far-applied toxicological approaches and indicate the need to reappraise the current mindset of toxicology and the claim to develop efficient, fast, and less expensive methodologies that will facilitate the drug safety investigations.

1.3. Animal based toxicology models

The animal-based models are the method of choice in drug toxicity/ efficacy investigations. The benefits of experimentation on animals are undoubted and have facilitated the progress of the medicine so far. The in-vivo experiments is a common requirement of all drug regulation agencies. However, the animal-based toxicology studies are slow, expensive, and present significant uncertainty [3]. The observed differences regarding the effectiveness/ toxicity of a chemical in animals and in humans raise questions about the animals-to-humans-translational success [4]. According to a recent review article, the animals-to-humans-translational success rate was 0-100%, regarding the toxicity studies, showing lack of model's reproducibility [4]. The testing of >2000 drugs in animal models showed that the obtained results / responses were inconsistent compared to those expressed in humans and, acclaimed that the successful cases were rather random. Thus, the animal models cannot provide consistent proof for the toxicity/ efficacy of a chemical at the human [3]. This inconsistency in animal-to-human-translation is explained by the fact that they are organisms, with significant physiological, genetic, epidemiological, biological, and biochemical differences. Also, the animal-models cannot predict late adverse effects, potentially occurring even years after the use of the drug. The limitations of the animal-to-human translation leads to wrong decisions, i.e., consider toxic drugs as safe, or consider safe drugs as "toxic", with serious impact on the volunteers' health [3].

1.4. New trends on toxicology

From the beginning of the century, the scientific community posed the constraints of the applied methods and focused on the development of novel, animal free models that are more human-relevant. Toxicology testing in 21st century, Tox21, is a unique collaboration between the federal agencies:

- National Institute of Environmental Health Sciences (NIEHS)/National Toxicology Program (NTP), National Institutes of Health (NIH), U.S. Department of Health and Human Services
- National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH)
- Centre for Computational Toxicology and Exposure (CCTE), Office of Research and Development, U.S. Environmental Protection Agency (EPA)
- U.S. Food and Drug Administration (FDA), U.S. Department of Health and Human Services,

that aims to develop new, animal-free ways, to assess rapidly the positive/negative impact of chemicals on the human health [5]. Tox21 proposes the use of emerging scientific tools towards the toxicity prediction, as the biochemical- and the cell-based assays employing robotics, the utilization of 3-D models of human tissues and organs, the experimentation on simpler than mice models (zebrafish), and the usage of advanced computational methods for the analysis and the interpretation of the resulted experimental data, focusing on the development of prediction models [5]. These tools are considered as more consistent in toxicity prediction, in contrast to animal-based models, as the results are measurable, reproducible, with defined uncertainty. So far, these efforts have achieved to investigate 10 thousand chemical compounds (drugs, food additives, industrial products, etc.), using tests that cover 125 biological processes. Moreover, they have detected 2800 genes in cells and tissues, that are useful in investigating the responses to toxic chemicals. The major aim of Tox21 is the detection of specific biological pathways, that their disturbance potentially leads to toxicity or ADEs [5].

1.4.1. *In silico* toxicology

In silico toxicology is an emerging field in drug safety surveillance, that utilizes computational resources (algorithms, databases, etc.) to predict the risk of toxicity induced by chemicals, as drugs, food additives, etc. The computational methods are deemed as complementary to the *in vitro* and *in vivo* experiments and their use facilitates the risk and efficacy assessment, minimizing the required cost and time. The *in silico* toxicity prediction models, developed by either experts or automatically - employing machine learning techniques, provide a major advantage: the ability to predict toxicity even before the chemical compound is synthesized, with only requirement that of the compound to exist virtually [6],[7]. The employed tools are data bases for storing data regarding the toxicity and the properties of the chemicals, software for the development of molecular descriptors, tools for the simulation of biology and molecular dynamics, methods to model the toxicity mechanisms and models for the toxicity prediction. In other words, *in silico* toxicology considers the chemical/ structural properties of a compound, to predict possible interactions with targets as proteins, genes, etc., in conjugation to recorded *in-vitro* and *in-vivo* evidence, aiming to virtually simulate the mechanism of the toxic response in cellular, tissue and even the organ level. These models interpret parameters as the genetic variations and the speed of metabolism [7], and the developed models can exhibit high consistency to the human organism. The workflow of a prediction model includes:

- Recording of biological data that provide information for associations between chemicals and toxicity end-points,
- Creation of molecular descriptors
- Development and evaluation of a prediction model
- Interpretation of the model [6], Figure 1.

Examples of the methodologies employed for toxicity modeling are [6]:

- Structural Alerts (SA) that are chemical structures that could induce toxicity directly or indirectly. They are often used in rules (rule-based models), defining that if a specific structure is presented, then a specific type of ADE is certain.

- Quantitive Structure-Activity relationship (QSAR): use of molecular descriptors to predict toxicity. These models are based on a regression-based idea and attempt to link a set of predicting variables, as the physicochemical properties with a biological response. Compounds that are described by the same QSAR exhibit high probability to express common mechanism.
- Dose- and time response models: describing relationships between the dose or the time with a defined biological effect.

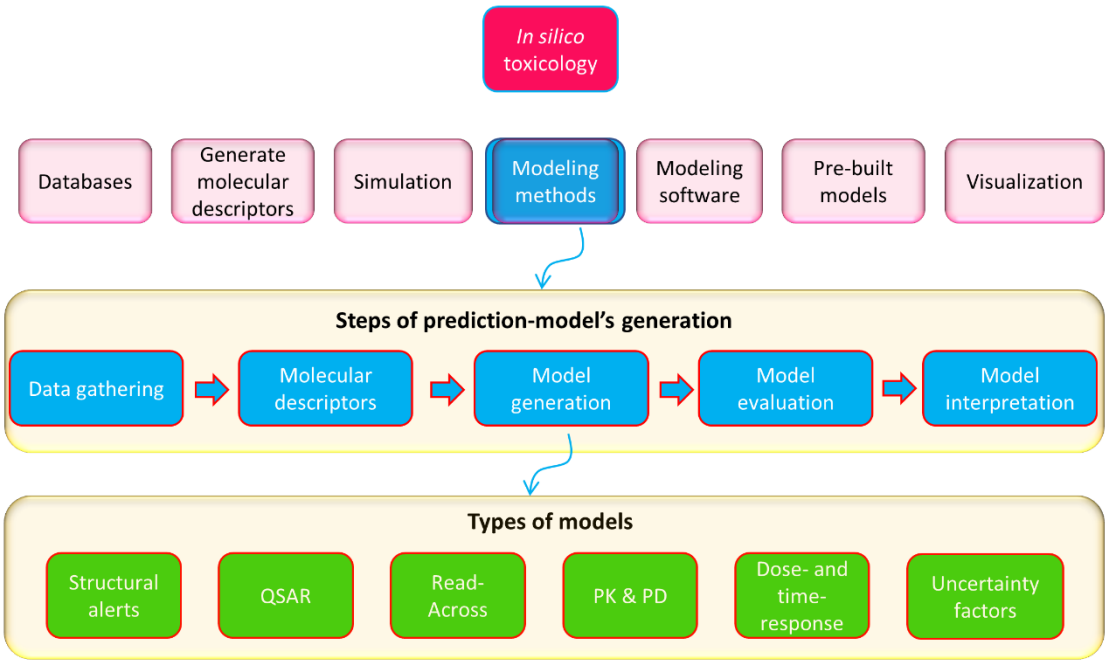

Figure 1 Graphical description of in-silico toxicology procedure [6].

1.4.2. 3D-cell cultures

Two dimensional (2D)-cell cultures have been used in the field of toxicology for many decades, allowing the investigation of chemicals' impact at cell-level *in vitro*. The method is based on the growth of cells on a flat surface, and so far, represents a key step of the drug development pipeline, as it is used for compounds screening before the *in vivo* testing. Despite the definitive contribution of 2D-cell cultures in drug safety assessment, the method's efficiency is considered as constrained by the inadequacy to represent with the

required accuracy of more complex cell environments, as the tissues. This gap was bridged by the use of animal models, but these models are also exhibiting limitations, as was discussed above. The recent years the implementation of three dimensional (3D)- cell cultures is proposed as an emerging, and better fitting to the purpose solution. The 3D-cell cultures are easy to use, versatile, quicker, less expensive, and more reproducible that the animal models [8], permitting more coherent observations. In contrast to the static environment of their 2D counterparts, the 3D-cell cultures provide a more comparable to the *in vivo* environment. I.e., the drug screening in 3D-cell cultures allows the determination of the induced biochemical response, considering the alteration provoked in the cell, and taking into account the interactions between the cells, the interaction of the cell and the extracellular cell matrix and the interaction of the cell with the whole organ as well [9]. The factors that impact the cell phenotype are the extracellular matrix composition (EMC), the matrix stiffness, the concentration gradients (oxygen, pH, nutrients, cellular metabolites, etc.) and, the stromal cells [8]. The components of ECM, i.e., the matrix proteins, the glycoproteins, the ECM sequestered, and the hepatocyte growth factors, etc., affect the cellular behavior in a tissue specific way, regulating the cell proliferation, differentiation, migration, the cell signaling and the cell death [8]. The biochemical composition of the ECM offers characteristic physical and mechanical properties, as the tissue stiffness which is essential for the homeostasis [10]. Thus, the simulation of the tissue in 3D-cell culture-model provides an accurate representation of cell's phenotype and, facilitates the comprehension of the biochemical processes involved, leading to the specific toxic end points.

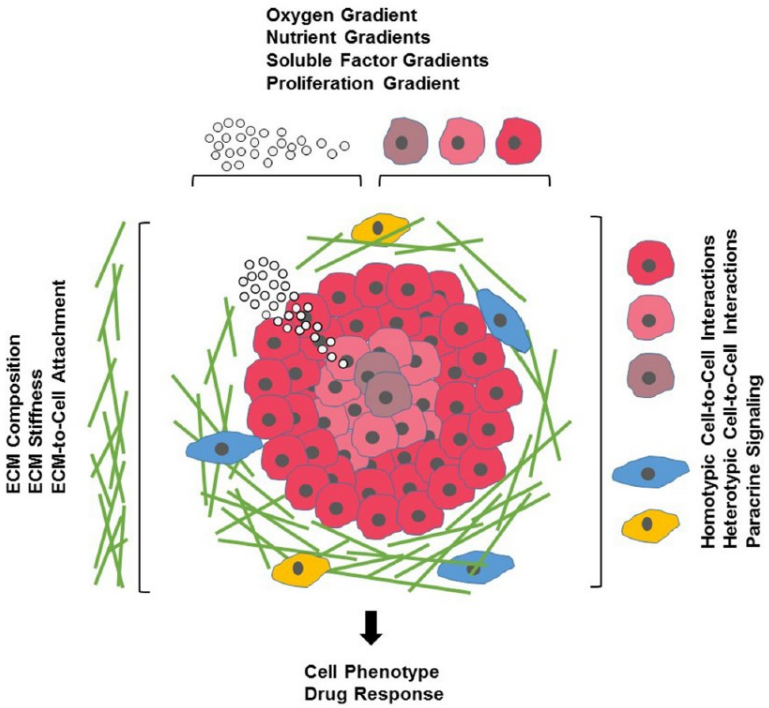

Figure 2 Graphical description of the complex cell environment [8]**.**

1.4.3. Systems biology & Omics

As described above, humans and animals are organisms of high complexity regarding their level of biological organization. This complexity leads to limited similarity between different species, i.e., humans and laboratory animals, and also limited uniformity in the individuals of the same species. Although humans share many common characteristics (biological, genetic, epidimiological, and biochemical), even the organisms of the twin brothers present differences. This dissimilarity leads to high uncertainty, even in human-to-human translation, posing barriers in the drug safety assessment, as the deep understanding of the pathways leading to specific toxic endpoint is challenging procedure. Drug induced hepatoxicity is a characteristic example, as it is a clinical endpoint of mitochondrial dysfunction; however, mitochondrial dysfunction results by many biochemical trailheads, e.g., interaction with multiple protein targets, fatty acid metabolism dysregulation, etc. Thus, the fact that two different organisms expressed a common toxic endpoint, e.g., hepatotoxicity, after their exposure to a specific chemical, cannot lead to the assumption that the chemical triggered the same biochemical pathways in them. The raising question is: *How can we predict toxicity without knowing the occurring biochemical alterations?*

Considering the timeline of the toxicity event, the compound firstly affects the cellular level in a direct or indirect manner, leading to e.g. stress response or apoptosis induction; then the cellular distribution, if expanded, leads to tissue or organ dysfunction, e.g. inflammation or necrosis; and finally is expressed as a clinical symptom, e.g., fever, organ-failure, etc. [11]. This toxicity timeline is probably affected by multiple molecular initiating events. Thus, it is important to understand in a more comprehensive way, to isolate and to define the biochemical background of the toxic responses in order to be able to predict them.

This attempt is facilitated by the emerging field of systems biology (https://irp.nih.gov/catalyst/19/6/systems-biology-as-defined-by-nih), that efforts to the establishment of interactions occurring among the different levels of biological organization, i.e., cell, tissue, organ, in order to provide more comprehensive insights into the organism's procedures. According to the National Institutes of Health, NIH, system biology combines the advanced computational models with the cell biology, the immunology, and the omics techniques (genomics, metabolomics, proteomics) to conduct general assumptions on the biology of complex species. According to Pius et. al., the application of systems biology in the toxicology emphasizes in:

- Absorption and distribution of the xenobiotic within the biological system.
- Transformation of the xenobiotic resulting in the generation of toxic and/or non-toxic intermediates.
- Interaction of the parent xenobiotic and/or its products with the cellular targets of toxicity.
- Alterations in the cellular molecules including genes, proteins, and lipids.
- Structural manifestations of target organ toxicity, for example histological changes.
- Functional manifestations of toxicity, for example impairment of critical functions in target organ(s)/organ system(s), and
- Elimination of the xenobiotic and/or the transformation intermediates from the biological system [12].

According to the central dogma of biology, the DNA (studied by genomics) is transcribed in RNA (studied by transcriptomics) which is translated into proteins (studied by proteomics). Finally, the actions of proteins result in the production of thousands of metabolites (studied by metabolomics) [13], Figure 3. The field of systems biology that focuses on the drug safety assessment, also known as systems pharmacology, so far, is based on toxicogenomics, a field that study the transcriptomics and proteomics interactions for the determination of the mechanisms that lead to ADEs [14].

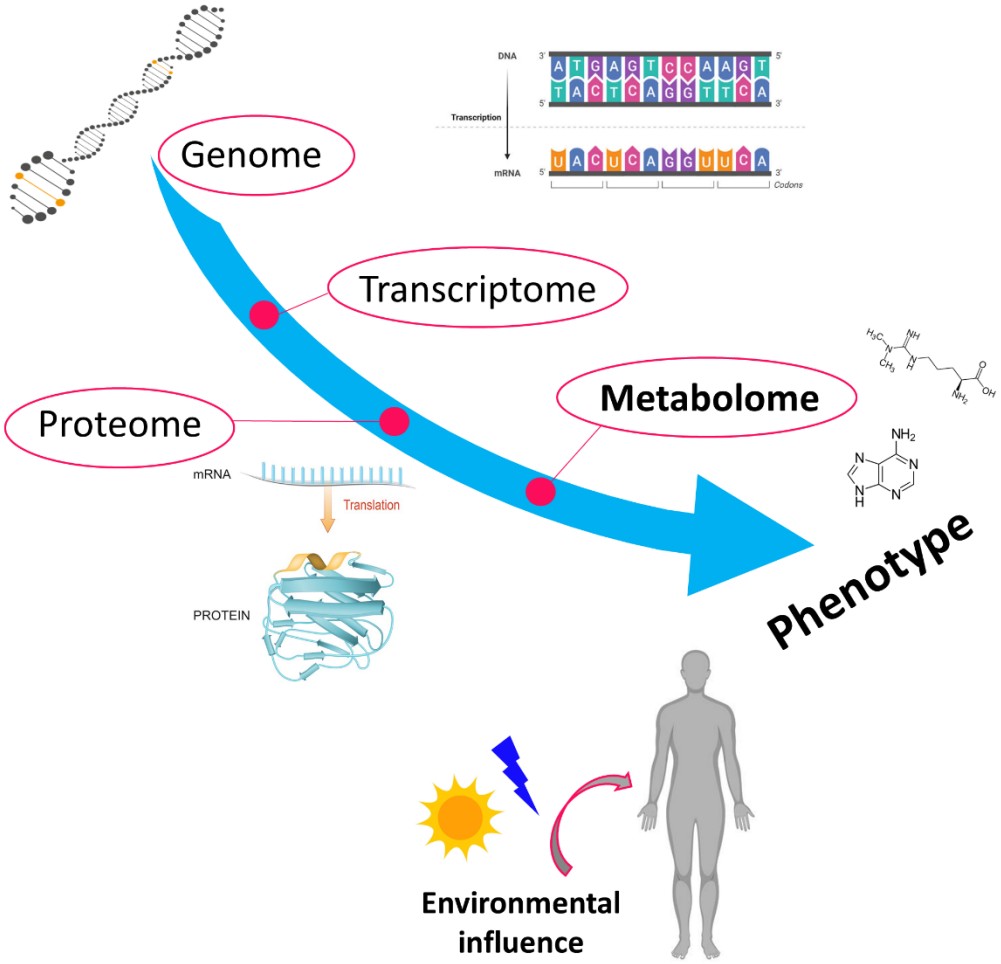

Figure 3 Factors that impact the profile of the phenotype, Figure modification of Steuer et. al.[15]

Transcriptomics exhibits extended application in the field of systems biology. From the 2002, 703 research articles that implement transcriptomics in the investigation of the drug toxicity and the drug safety, have been published (PubChem: transcriptomics drug toxicity & transcriptomics drug safety). This

field encompasses the emerging analytical techniques to detect alterations in genes' expression regulation, when an organism is exposed to specific xenobiotics, i.e., drugs, chemicals, environmental pollutant [12].

The investigation and development of protein-protein interaction (PPI) via proteomics, is another popular approach in the field, and combined with the transcriptomics form the field of toxicogenomics, which attempts to relate the gene and the proteins expression provoked by external stimuli.

Despite the pronounced benefits of toxicogenomics, there are limitations that should be addressed. Transcriptomics studies are topological limited, in other words, specific tissues are submitted in transcriptomics analysis during a toxicology study. However, so far, the scientific community has not set criteria or rules for the tissue selection and the appropriate time of the sampling, considering that they are slow or fast, and strong or weak responders. Commonly, the interest is focused to the leucocytes and the examination of the liver, as it is the organ that is mainly involved in drugs metabolism. Therefore, the results obtained from liver's investigation are permanent for the renal or the cardiac function as well. A holistic picture would be formed if the transcriptomics and proteomics analysis were used to screen all tissues, however this would be very expensive. Moreover, in transcriptomics it is challenging to isolate the relevant-to-the-toxicity response from the adaptive or the homeostatic response that also occurs. The major constrains of toxicogenomics, in the effort to understand and predict the toxic response of a patient to the drug, is that other contextual factors, i.e., age, lifestyle, diet, health, stress, polypharmacy, microbiota, etc., are not considered [16]. On the other hand, proteomics is more informative regarding the contextual impact factors, although they are not informative for the enzymatic activities related to drug toxic response. However, several toxicities are related to alteration of enzymes function, and these alterations are not always reflected to protein concentration per se [17].

Metabolomics deemed to be promising in systems biology, as metabolites represent the final products of multiple interactions between genes, transcripts, and proteins. The temporal metabolic profile is sensitive to all external and internal stimuli of an organism, and provide information for the organism

responses, not in an isolated manner, but as results of actions exhibiting in parallel in a complex system[18]. Metabolomics are fast and cost-effective, widely applied in biofluids analysis and they can be repeated in an effort to identify the fast and the slow responders [17]. The top-down and the bottom-up data reduction approaches that are commonly used in system biology studies are described in Figure 4.

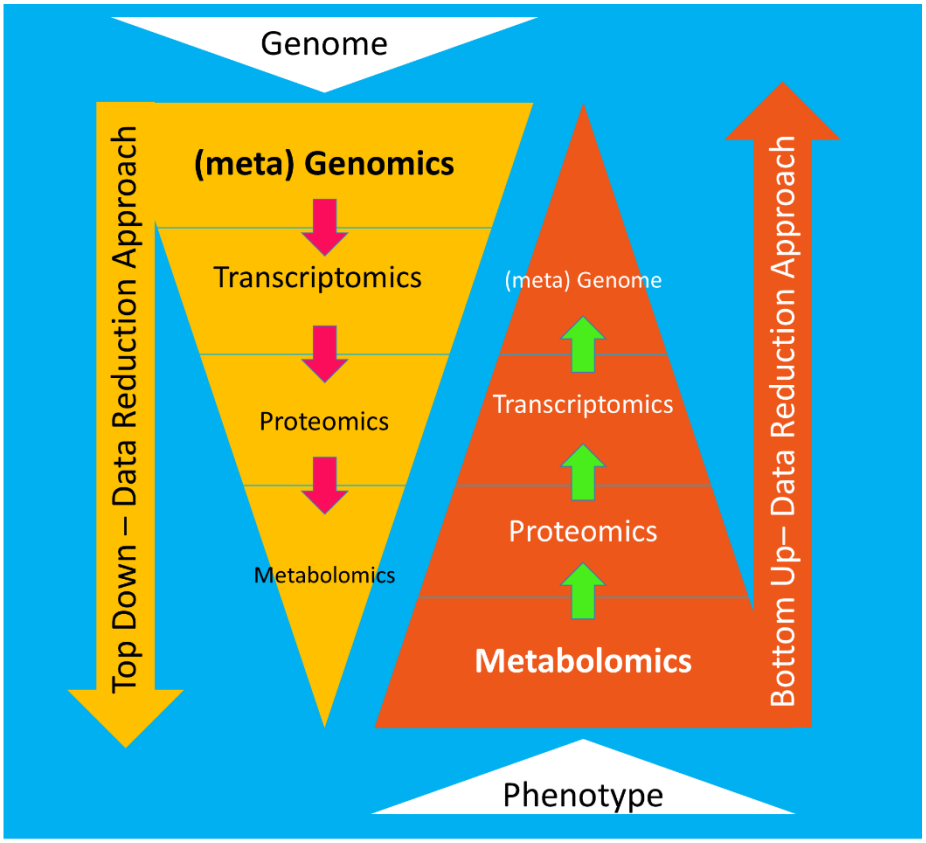

Figure 4 Top-down and bottom-up data reduction approaches in system biology, Figure modified from Pinu et.al [18].

1.4.4. Points of toxicity

Systems biology uses emerging post-analysis methodologies towards the development of biological networks, combining the evidence resulting from independent omics datasets. It is essential that the omics data are biochemical evidence linked to alterations that precede the pathological phenotype. Thus, due to this temporal disconnection between the omics and the clinical evidence,

it is challenging to connect clinical signs with the initiating pathway. However, the omics-based knowledge, merged with data-base clinical evidence, can facilitate the discovery of subclinical biomarkers. Under this notion, the Consortium on Metabonomic Toxicology, COMET, is a large project that started in 2005, aiming to determine shared drug-induced metabolic responses. The main idea is that drugs of different mechanism of action trigger specific biochemical pathways that can be used as the basis of an *in-vivo* screening process. If we could identify specific metabolic responses and pathways that are shared in toxic responses, we then would be able to screen them and predict the future toxicity even *in-vivo*. Thus, COMET conducted a large number of drug innervational and physiological experiments to make a data-base of metabolic responses data and combine them with conventional clinical data [17].

2. Chapter 2

Metabolomics as a powerful tool in drug safety/efficacy studies

2.1. Metabolomics in Toxicology

Metabolomics is the field of analytical chemistry that studies metabolome. The term “metabolome” refers to the total amount of low-molecular weight compounds (<1500 Da) existing within the biological systems, and to their interactions. The composition of the metabolome is defined by the genome, transcriptome, and the proteome, but also by other factors, i.e., environmental, lifestyle, drugs, underlying diseases, etc. Human Metabolome Data Base, HMDB, version 5.0 includes 220,945 metabolites entries (https://hmdb.ca/, 2023). However, the total number of human metabolites is not known so far, due to the high complexity of the metabolome. The metabolites appear in the human organism at concentration scales that range from nM to mM. Metabolites can be categorized as (i) intracellular (endometabolome), including the metabolites produced by cells, tissues or organisms; (ii) extracellular (exometabolome), including metabolites secreted or consumed by cells; (iii) microbial, produced by microbiota and; (iv) xenomatabolome, formed by metabolites that are produced by xenobiotics, pollutants, and nutrition [13].

The majority of the community uses the terms metabolomics and metabonomics as equal, however, they deem to have specific scientific aims: metabolomics efforts towards the identification and qualification of all, both endogenous and exogenous, metabolites existing in a specific biological sample. On the other hand, metabonomics focuses on the alterations occurring in the metabolic profile as a consequence of a specific stimuli, i.e., pathological condition [13]. In the current text, these terms will be used interchangeably.

2.2. Metabolomics pros and cons

As metabolites’ composition is affected, in a high-sensitive-manner, by both the steady physiological equilibrium of organism and, the changes in the environmental stimuli, the metabolome is considered as an accurate predictor for the phenotype, and the metabolomics deemed to be the most functional of the omics approaches [19],[20]. An important benefit of metabolomics is that it facilitates the animal-to-human-translation, or in general the translatability from

experimental models to humans, as metabolites have universal chemical structure, despite the organism they belong. Regarding the limitations of metabolomics, the most severe is the existence of multiple confounders inducing alterations in the metabolic profile, making it challenging to identify latent consistent correlation of a metabolic change by a specific stimulus. Also, the lack of direct association with the genome, which constrains the interpretation of the metabolomics results possess another major issue.

2.3. Metabolomics views and strategies

From the analytical chemistry scope, there are three main approaches: (i) the global, or untargeted metabolomics, (ii) the metabolic profiling and, (iii) targeted metabolomics[13]. Global metabolomics investigates the total amount of metabolites that could be detected by the instrumentation employed, i.e., QTOF. This approach is mainly selected in differential studies, i.e., studies focusing on the metabolic differences of two groups, where there is no prior knowledge concerning the metabolites that will be proved as informative. Therefore, total screening of compounds is employed, resulting in a vast amount of data, that should be subsequently interpreted using advanced bioinformatics tools. The signals considered as discriminating are identified, in a statistical driven identification procedure. On the other hand, metabolic profiling is based on the quantitation of predefined metabolites that belong to the same class, i.e., amino acids, or participate in a specific biochemical pathway [13]. Finally, the targeted approach is based on a prior hypothesis and investigates the regulation of specific metabolites, designing appropriately the experiment, and the analysis conditions [21].

Moreover, regarding the differential studies, we can distinguish two subcategories, the metabolic fingerprint, and the metabolomics footprint. The first refers to the determination of the intracellular metabolome, and the second focuses on the extracellular metabolites that are excreted or consumed by the cells [13].

2.3.1. Pharmacometabolomics

Pharmacometabolomics (PMx) is the field of metabolomics that efforts the assessment of drugs' efficacy or toxicity. PMx studies attempt to detect specific biomarkers or metabolic patterns that facilitate the comprehension of drugs

action, metabolism, ADEs, etc., and their ultimate goal is the development of models that can predict the individual's response to the drug [16]. In the effort to estimate the drug induced effects, there are two main design approaches that are followed, based on the intended conclusions. The first idea is the comparison of the baseline metabotype, corresponding to pre-treatment samples, with the treatment-metabotype, corresponding to after- or during-treatment samples. This comparison describes the metabolic alterations induced by the drug. The second idea in the comparison of the baseline metabotype, with the longitudinal drug exposure in order to estimate the dose-response [16].

2.4. Challenges on metabolomics design

The ultimate goal of metabolomics applied in drug safety/ efficacy studies is the identification of sensitive biomarkers that enable the early prediction of the organism's response to the studied drug. The effort to determine metabolic biomarkers that could be used in the clinical routine is constrained by several challenges that impact the metabolomics analysis procedure. A major issue is the chemical diversity of the metabolites, and the existence of isobaric and isomer compounds that hinder the identification. Also, the existing chemical diversity of the metabolome requires the combination of different analytical techniques to obtain the holistic view, such as gas and liquid chromatography, different sources of ionization, different chromatographic columns, as it will be further discussed below. Moreover, there is a small number of available authenticated standards of metabolites, i.e., 600 reference standard of IROA MSMLS, Biocrates, etc., and thus it is challenging to validate other potential biomarkers that might be more informative than the known amino acids, fatty acids, etc.

In addition, as metabolome is affected by multiple sources of variation, is deemed essential to design proper protocols to reduce the preanalytical and the batch-to-batch variation, as their existence will miss-lead the metabolomics statistical interpretation.

Another key point in the metabolomics design is the estimation of the proper sample size, especially in cases of clinical cohorts. It is suggested that the

impact of non-relevant to the case variation has serious impact when studying small cohorts, i.e., 20 persons, and the variation is decreased when the sample size in large, i.e., hundreds of individuals, as it is more representative for the total sample population [20]. The thorough design of the size and the quality of the cohort is essential to obtain accurate and consistent results. However, this design cannot overcome the impact of the demographic's imposed variability, that should be addressed separately, during the statistical interpretation of the omics data. The Husermet project is an effort to analyze the metabolome of the human serum. Part of this project is the determination of the manner that the sample size and the demographics affect the phenotyping quality [22]. So far there is well established knowledge on the impact of the sex, the age and the BMI on the expression of clinical conditions as the kidney disease / failure, the fatty acid oxidation disorders and others [20].

2.5. Metabolomics and mass spectrometry: Applications & Trends

As mentioned above, the metabolome is the sum small molecules (<1500 Da) that exist in a biological system, that exist in a wide range of concentration levels. Metabolome includes compounds with wide chemical diversity i.e., lipids, sugars, amino acids, small peptides, steroids, etc., and therefore it is impossible to investigate the whole metabolome using only one analytical technique. The most common approach is the combination of orthogonal separation techniques coupled to detectors. The most commonly implemented platforms for metabolomics analysis are the nuclear magnetic resonance (NMR) and the liquid chromatography tandem to mass spectrometry (LC-MS) methodologies, however, gas chromatography coupled to mass spectrometry is also used.

NMR belongs to the spectroscopic techniques and the method investigates the energetic transition of nuclear spins in the presence of a strong magnetic field. NMR has been a useful tool in life sciences and especially in the field of metabolomics. The stronger advantage of NMR is that facilitates the identification and the structure elucidation of the organic compounds. As NMR measurements are sensitive to the chemical environment, the technique provides selective information about the compounds regarding their physiological characteristics [23]. NMR is a versatile technique, as it is

nondestructive, nonbiased, facilitates the qualification and requires minimum sample preparation. Moreover, NMR is highly reproducible and enables the facile determination of sugars, organic acids, alcohols, etc., that are challenging for LC-MS platforms to detect [24]. The major limitation of NMR is its low sensitivity, compared to other methodologies, as it provides detection limits that range from low-micromolar to high-nanomolar values. This constrains the use of NMR in biomarkers detection, as it cannot detect low abundance metabolites that may be important as differentiating variables. It is worth noting that NMR permits the reliable detection of a few hundred, while it is known that human metabolome includes several thousands of metabolites [24].

2.5.1. HRMS

High resolution mass spectrometry (HRMS)-based platforms provide an efficient alternative in omics and particularly in metabolomics. In contrast to NMR, HRMS methodologies permit full-scan acquisition, offering a more comprehensive view on the metabolome. They offer wide detection coverage, and thus high-throughput metabolomics profiling, combined with high sensitivity, precision, and accuracy. To the HRMS class belong mass spectrometric analyzers that enable the detection of analytes to the nearest 0.001 atomic mass units, i.e., time of flight (TOF), orbitrap, and Fourier transform ion cyclotron resonance (FTICR) analyzers. In general, the mass spectrometer measures the mass-to-charge ratio (m/z) of one or more charged compounds. In contrast to the mass spectrometers as the triple quadrupoles, the HRMS instruments permit the acquisition of multiple scans and the ability to record all the charged molecules that exist in the sample. This characteristic is crucial in omics studies, therefore HRMS-based platforms are the most commonly viewed in the literature. In contrast to TOF, FTICR and orbitraps provide higher resolution and mass accuracy, however, their cost seems to constrain their use. On the other hand, TOF gains place as a versatile instrumentation of high consistency and especially high acquisition speed.

2.5.2. UPLC-ESI-QTOF

The liquid chromatography-mass spectrometry (LC-MS), and particularly ultra-performance liquid chromatography (UPLC), has become one of the most commonly used techniques for metabolomics studies, offering high sensitivity,

selectivity, and throughput in data acquisition. As mentioned before, the metabolome includes compounds exhibiting high chemical diversity which is expressed in their physicochemical properties. Thus, the use of orthogonal separation techniques is a common practice in metabolomics studies. The most common approach is the conjugation of complementary types of LC, i.e., combination of hydrophilic interaction liquid chromatography (HILIC) with reversed phase liquid chromatography (RPLC), or the combination of positive and negative ionization modes.

HILIC columns have a highly hydrophilic stationary phase, e.g., amides, providing the retention of polar metabolites, i.e., amino acids, nucleic acids, the sugars, and small organic acids. HILIC is considered as complementary to RPLC which is the most commonly applied. RPLC columns have a lipophilic stationary phase for the retention of non-polar lipophilic molecules, as the lipids and the fatty acids. The combination of these to LC approaches provides more comprehensive metabolites coverage, as it facilities the better performance of both polar and non-polar compounds [25].

LC and MS platforms are coupled together via an ion source, that transforms the LC-eluted analytes into charged ions that can be further analysed by MS based on their m/z ratio. The ion sources are usually described as *soft* or *hard*, depending on the fragmentation occurring during analytes' ionization. Soft ionization methodologies are considered as more appropriate for metabolomics workflows, as the metabolites can be charged in their intact form, and thus, it is easier to identify them. The electrospray ionization source (ESI) belongs to the family of *soft* ion sources and is the most famous technique in metabolomics. ESI generates mainly single charged [M+H]+ or [M-H]- ions, and also a variety of other adducts, e.g., [M+K]+, [M+Na]+, [M+NH_4]+, [M-H_2O]- [M+Cl]-, etc. During the ionization procedure, the eluent exits from the column and passes through a steel needle under the impact of a high voltage, ranging from 2-6 kV, resulting in the generation of charged droplets. These droplets, containing ions of analytes and solvent, insert the cone, Figure 5, where they are heated and evaporated under a strong electric field that leads to the formation of droplets with increased Coulombic interactions on their surface. These droplets sequentially decompose into smaller ones until the analyte is released from the

solvent, passing in the gas phase. The analytes are transferred into MS analyzers usually using with dry N_2 [26].

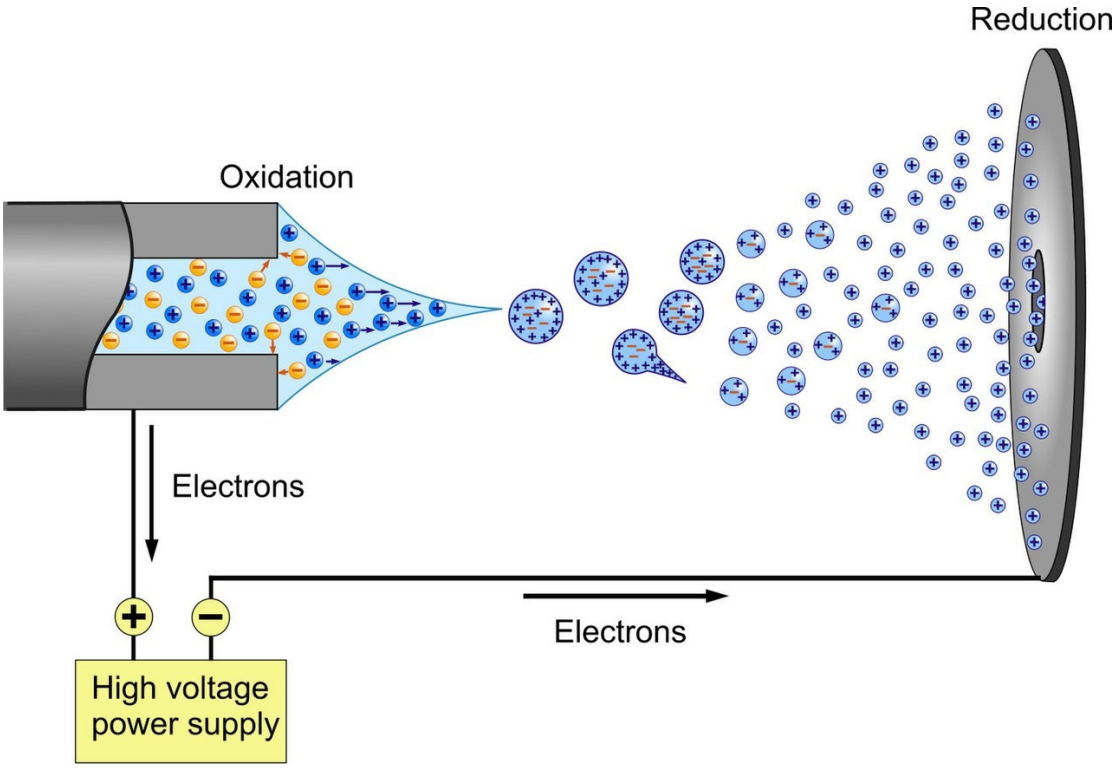

Figure 5 Graphical description of electrospray ionization (https://en.wikipedia.org/wiki/Electrospray_ionization#/media/File:ESI_positive_mode_(21589986840).jpg)

The QTOF mass analyzer is a promising 'hybrid 'analytical platform that combines the advantages of a quadrupole (Q) and a TOF analyzer. Quadrupoles offer increased efficiency in compounds fragmentation, i.e., highly selective fragmentation. TOF analyzers permit rapid acquisition speed and high-resolution capability. This combination enables the full MS scan with the ability to fragment specific compounds and increase the identification confidence, and full scan of fragments as well. TOF mass analyzer provide the separation of charged ions during their 'journey' into a flight tube under high vacuum. The ion separation is based on m/z ratio. The charged ions are gathered in the area of the orthogonal accelerator, and are subjected to a homogeneous electric field pulse, providing them the with equal dynamic energy. The ions are 'pushed' in the TOF tube, a region free of field- and drift impact. The ions start their rout with equal kinetic energy, due to the conservation of the energy, and they acquire velocity that is inversely

proportional to their mass [27]. At the end of TOF tube the detector records the signal of incident ions. A scheme of a o-QTOF system is presented in Figure 6.

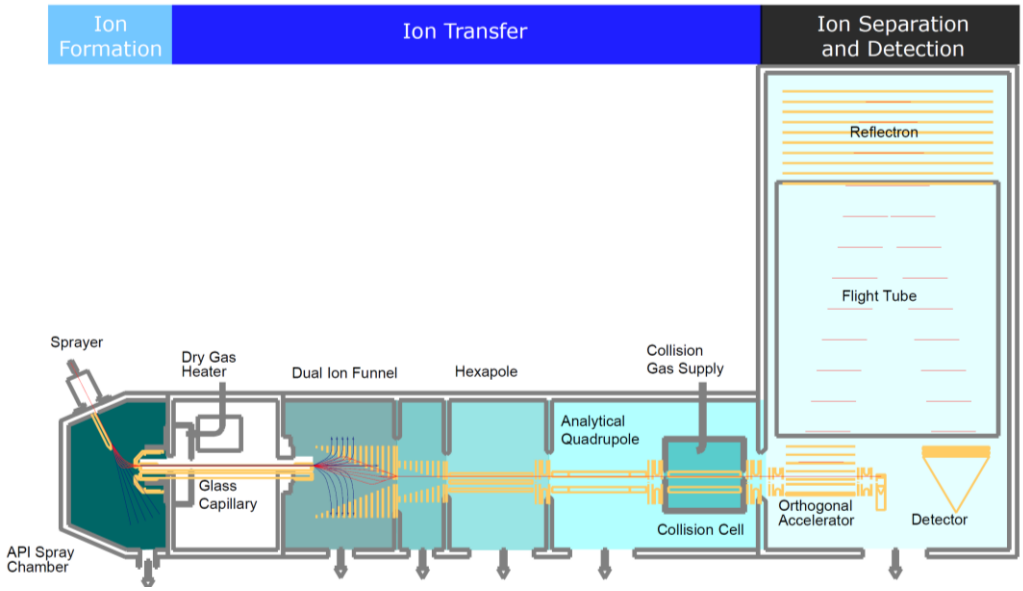

Figure 6 Graphical derscription of the QTOF system (https://www.researchgate.net/figure/Schematic-of-the-impact-II-mass-spectrometer-not-to-scale_fig1_277085139)

The first part of the Bruker Impact or TIMS-TOFpro2 LC/o-QTOF platform is the ionization source, that connects the LC system with the MS. The sample exits the LC system, enters the source and charges are developed. Gradually the mobile phase is dried and heated through a glass capillary in order to remove the solvent. The charged ions are collected with high efficiency and focused via the dual ion funnels and the hexapole and subsequently directed to the analytical quadrupole (Q1). The Q1 is followed by the collision cell (Q2), and the orthogonal accelerator. The orthogonal accelerator, also known as the pusher-puller system provides the same kinetic energy to the ions and pushes them through the time-of-flight tube. QTOF is mostly applied for 3 alternative spectra acquisition modes:

i) full-scan MS: Q1 and Q2 are set to the broadband mode, giving free access of all ions in the TOF tube;

ii) full-scan MSMS, or full-scan of the fragments: Q1 is set to the broadband mode and Q2 is on, conducting fragmentation using accelerated flow of neutral gas, e.g., N_2, and high voltage. The

produced fragments are accelerated through the pusher puller and enter the TOF;

iii) Full-scan MS and selective MSMS: the precursor ions enter the TOF tube unhindered by the Q1 or the collision cell. Then in the following scan-rounds, specific precursors are isolated by the Q1, applying the appropriate RF conditions, and these precursors are submitted to fragmentation and their produces fragments enter the TOF, providing information of selective fragmentation [28].

2.5.3 DDA, DIA and SWATH acquisition strategies

LC-MS platforms as the UPLC-QTOF, permit the performance of both targeted and untargeted (often called as non-targeted or non-target) metabolomics studies. In targeted approaches the data acquisition focuses on a pre-selected set of metabolites, based on a prior hypothesis, and therefore employ techniques as the multiple reaction monitoring (MRM) and the parallel reaction monitoring (PRM). On the other hand, untargeted metabolomics require comprehensive detection of all measurable metabolites existing in the samples.

The most applied LC-MS-based techniques in metabolomics profiling is the data dependent acquisition (DDA) and the data independent acquisition (DIA). DDA and DIA use different strategies to conduct simultaneous acquisition of MS1 (precursor ions) and MS2 (fragments) data of all the charged molecules existing in a sample. In both cases, there is a data acquisition cycle that contains the acquisition of a full MS1 scan, followed by the acquisition of one or multiple MS scans.

In DDA, the precursors detected in the MS1 scan are ranked by their abundance, and the top n of them are isolated for fragmentation, to receive their corresponding MS2 one by one. The major advantage of DDA is that the fragments are inherently assigned to one specific precursor, and this information facilitates the identification and increases its confidence. The constraints of DDA is that provides low coverage of MS2, as only the more abundant precursors are subjected to fragmentation, and also the MS2 spectra quality is undefined. This happens because the MS/MS spectra are not always acquired at the apex of the chromatographic peak [29].

In contrary for DIA, all precursor ions are fragmented providing multiplex MS2 data. In this way DIA permits the acquisition of MS2 for all precursors. In order to perform the fragmentation, the precursors are sequentially isolated in a predefined m/z window ranging from 5 Da to a full mass range, and then they are fragmented in one or more repeated fragmentation cycles. The main challenge of DIA is the lack of direct link between MS1 and MS2. The width of the isolation window is the differentiating point of the several DIA methods, as the all-ion fragmentation (AIF) and the SWATH. In AIF, also known as MS^E, all the co-eluted precursor ions of the whole mass range are transmitted for fragmentation. In this case, the product ions are assigned to precursors by the retention time, the mass defect, the peak shape similarity, or by combination of them. Nevertheless, the complexity of the MS2 spectra is increased, and thus new approaches that address this complexity seem to gain place. SWATH is a DIA technique that enables the sequential fragmentation of all precursors in a serial of quadrupole (Q1) isolation window, that typically ranges from 5-100 Da. In this way, lower number of precursors is simultaneously fragmented, and thus the complexity of MS2 is decreased [29]. DDA and DIA data acquisition are graphically represented in Figure 7.

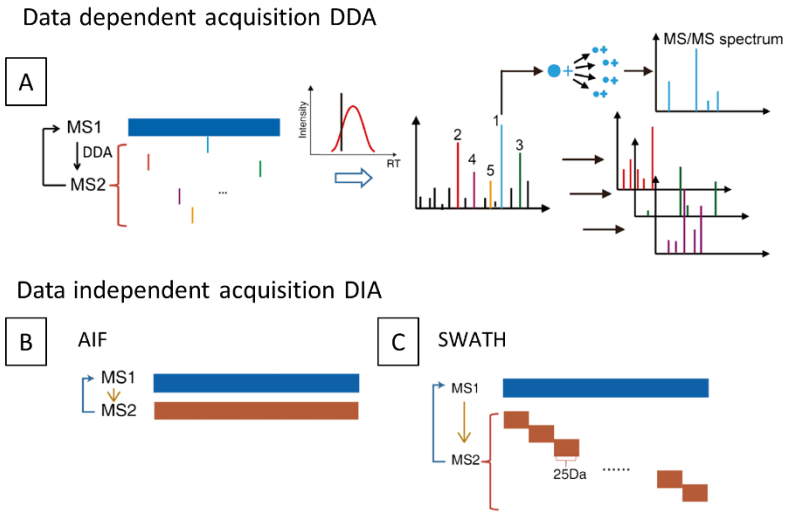

Figure 7 Graphical representation of data acquisition methodologies: A) Data dependent acquisition; B) All ion fragmentation (type of data independent acquisition); SWATH (type of data independent acquisition). Figure modification from Wang et. al. [29]

2.6. Metabolomics identification

The identification, namely the attribution of a specific Accurate Mass Retention Time (AMRT, mz_rt) feature to a known metabolite, is a challenging procedure. In the case of DDA, the structure and the metabolite can be clarified by comparing the similarity of m/z of precursor and its MS/MS spectra to the metabolites standards existing in the spectral libraries MS1 and their fragments can be directly searched in databases such as the HMDB [30], the METLIN [31], the MassBank [32] and numerous others. In the case of DIA, the identification requires one more important step: the creation of pseudo-MSMS that are assigned to a specific precursor.

The generation of spectra clusters, either in MS1 or MS2 levels is facilitated by several algorithms. The CAMERA algorithm generates MS1 clusters employing peak tables, based on the retention time, the isotopes, the peak shape, and the adducts[33]. The MSClust algorithm performs clustering of ion-fragments in the dataset that originate from a single metabolite, i.e., it is suitable for DDA datasets, based on two properties: (i) similarity of chromatography, especially focusing on the retention time, and (ii) quantitative similarity of ion-fragment patterns across a number of samples analyzed [34]. The RAMClust algorithm uses high collision energy MS without precursor ion selection, acquired simultaneously with the low collision energy MS. Thus, it is suitable for DIA data. RAMClust is based on the assumption that two features, produced from the same compound, present similarity in their retention time and also high correlation in their abundance profiles between different samples of the same dataset [35].

The identification strategies of metabolomics can be categorized as library-free and library-based. This discrimination is not related to the treated data (DIA or DDA) but to the existence of a prior hypothesis. Library-based approaches, also known as *screening of suspect compounds,* focus on the detection of specific metabolites, or metabolites belonging to a specific class, e.g., amino acids. These efforts are typically biologically driven, as the metabolomics study aims to investigate alterations occurring in a specific group of metabolites, or specific metabolic pathway. On the other hand, library-free methods are also hypothesis-free and are driven by the statistical interpretation. Namely, the all

the detected features are employed for the statistical comparisons, e.g., classification models, student t-test, etc., and those exhibiting the higher statistical importance are submitted to identification, using algorithms as mentioned above.

2.6.1. Dark metabolome and new trends on metabolomics identification

Untargeted metabolomics studies, especially the high-throughput MS-based, result in a vast number of mz_rt features. However, more than 98% of these features remain unknown. The major issue is the lack of MS/MS metabolites reference spectra. It is characteristic that PubChem includes more than 60 million compounds, but only 20 thousand of them are linked to MS/MS spectra [36]. The unknown features, also considered as dark metabolome, may be unregistered metabolites, or products of the metabolites' metabolism, etc. These features commonly bear increased statistical power in metabolomics studies, however, they cannot be considered as potential biomarkers as they cannot be characterized. Thus, the scientists of the field seek for alternative identification ways. The prediction of MS/MS spectra of given structures seems to be a powerful tool in spectral interpretation. The generation of predicted MS/MS is actually the prediction of potential fragments and intensities, based on the chemical bond breakage rules and the number of fragments. This effort of MS/MS prediction is also called *in-silico* fragmentation. MyCompoundID.org (MCID) is an online compound library that includes ~8 thousands metabolites, existing in HMDB, and >375 thousand predicted human metabolites in a developed Evidence-based Metabolome Library (EML). The EML is a library of predicted products resulting from metabolites metabolism, with one metabolic reaction. The *in-silico* fragmentation method employs heteroatom-initiated bond breakage rules and is applied in all MCID metabolites [37].

2.7. Metabolomics interpretation and biomarkers determination

The biomarkers or the biological markers represent a subgroup of medical signs. The medical signs differ from the medical symptoms, as they are *objective indications of medical state observed from outside the patient and can be measured accurately and reproducibly* [38]. According to WHO, biomarker is "any substance, structure, or process that can be measured in the body or its products and influence or predict the incident of outcome or disease".

According to their use, the biomarkers are categorized in groups:

The *diagnostic biomarkers* enable the detection or the confirmation of the existence of a pathological state or condition of interest.

The *monitoring biomarkers* are measured serially and allow the assessment of the status of a disease or medical condition as an evidence of organism's exposure to a drug, chemical or environmental agent.

The *pharmacodynamic or response biomarkers* are used as the alterations of their levels describe the response to exposure to a medical product, environmental agent, etc. These biomarkers are commonly used in clinical practice and early therapeutic development and also are useful in the drug dosing.

The *predictive biomarkers* are compounds that their existence or the alteration of their levels can identify a person or group of persons as more vulnerable to exhibit an effect (favorable of not), due to their exposure to a drug or an environmental agent.

The *prognostic biomarkers* are employed to identify the probability of a clinical incident, disease relapse, or disease progression in patients with a specific condition of interest. The prognostic biomarkers are used as criteria to classify the high-risk groups. The prognostic biomarkers are used for the risk assessment, i.e., for ADEs occurrence invoked by a specific disease, e.g., heart failure due to chronic kidney disease, while the predictive biomarkers describe the response of an individual to an external stimulus, as the drug administration.

The *safety biomarkers* are medical sings that their levels before and after an exposure to medication or to environmental agent suggest the probability of toxicity as a side effect.

The *risk biomarkers* describe the likelihood to express a disease or medical condition for a person who does not currently express clinical symptoms. The difference of the risk and the prognostic biomarkers is that the second investigates the vulnerability to express a pathological state, while a specific disease is already occurring [39],[40].

2.7.1. Biomarkers validation

According to PubChem, there have been published 25,959 original articles and only 1,441 patents regarding “biomarkers detection” between 2010 and 2023, suggesting that only a low percentage of the discovery studies end up in biomarkers of clinical use. This gap results from the high false discovery rate of biomarkers or by the *true* biomarkers with weak clinical performance, i.e., low specificity and sensitivity, etc. In order to render the knowledge obtained from the metabolomics actually applicable in the field of life science, the procedure of biomarkers’ discovery should include additional steps that ensure the consistency and the fit-of-purpose of the potential metabolite. Thus, according to Koulman et. al., the phase of the differential study that highlights differentiating metabolites, must be followed by two qualification phases and one application phase. The *qualification phase I* focuses on the identification of the metabolite, on the validation analysis, and on the evidence of biomarker’s importance, provided by a new clinical study. The final goal of this step is to decide if it is cost-effective to examine a differentiating metabolite as a potential biomarker. The *qualification phase II* is related to the validation process of analytical methodologies, as they efforts to provide experimental evidence regarding biomarker’ precision, reproducibility and, robustness and also record the limitations of its use by testing its efficiency in a wide cohort. The second step aims to answer weather the alterations of metabolite’s concentration level clearly indicate a specific pathophysiology when other symptoms are considered [41].

As the validation steps intermediate between the metabolite’s discovery and biomarker’s development require plenty of time and resources, it is deemed essential to develop thorough metabolomics studies that minimize the probability of false discovery, concerning the resulted differentiating metabolites. This attempt encompasses the identification of high confidence, as it was discussed above, and thorough variable selection pipelines that will be discusses below.

2.7.2. Biomarkers selection

The thorough variable selection process is deemed essential in naïve studies of biomarkers discovery, as it minimizes the probability of false discovery,

before the step of biomarker validation, saving time and resources. In differential metabolomics studies, i.e., comparison of control and case cohorts, the variable selection is mostly based on the combination of univariate methods, e.g., Student's t-test and/or receiver operating characteristic (ROC) analysis for features steming from multivariate statiststics as PLS, PLS-DA etc., Figure 8 [42]. The multivariate models are used for classification purposes and then their most discriminating variables are considered as important, for example variables with VIP scores > 1.5. However, the reliability of this strategy is disputed by the high false discovery rate occurring in the validation level. Thus, complementary variable selection methods should be included in metabolomics pipelines and should be selected based on the scope of the study and the characteristics of the cohort or of the experiment.

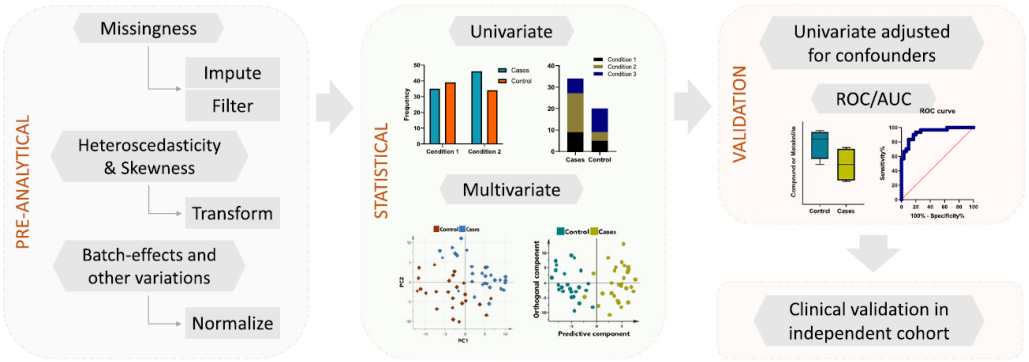

Figure 8 Simplified pipeline of the statistical procedure that is commonly applied in metabolomics studies [42]

2.7.2.1. Common strategies in metabolites selection

Multivariate analysis (MVA) is a powerful tool in the metabolomics field, enabling the statistical interpretation of the data considering their interactions as well. Biological systems show increased complexity as already discussed, and thus, the phenotype-differences of two cohorts, i.e., diseased, and healthy, cannot be easily represented by alterations in single variables. The study of system-level changes provides a more accurate view of the investigated phenotype, elucidates the multiplex impact of the condition, and facilitates the investigation of characteristic biomarkers. MVA interprets all variables in parallel, estimates putative correlations between them but also the provides semi-quantitative insights of the detected correlations to the phenotype of interest [42]. This is the main reason tat MVA methods find extended application

in the metabolomics field, in contrast to univariate strategies that are usually used as the final step of biomarkers statistical validation.

The MVA models are categorized as supervised and unsupervised. Unsupervised models use an impute of variables (X) and effort to find relations between objects, namely, to generate object clusters, or to find classification trends based on data reduction. Supervised models use both variables (X) and known information of the samples (Y), as the phenotype, aiming to generate (x,y) functions able to predict group classification, or predict a continous response. Principle component analysis (PCA) is the most popular unsupervised method, that detects independent components in the data, based on linear correlations of the correlated features. PCA is widely used in metabolomics, to highlight potential trends, however, serves little purpose in biomarkers discovery. PCA is mostly used for data evaluation, in order to identify hidden confounders, detect outliers, or to check the reproducibility of quality control (QC) samples.

On the contrary, unsupervised MVA exhibits critical role in biomarkers detection. Partial least square discriminant analysis (PLS-DA) has numerous applications in differential metabolomics studies, as it aims to identify the best predictor variable (X) that explains the response variable (Y). In other words, PLS based methods seek to reveal those features that impact the phenotype of interest. Orthogonal PLS-DA (OPLS-DA) is a promising alternative to PLS-DA, that aims to separate the existing variance into irrelevant (*orthogonal* variance) occurring by noise or confounders and into relevant that is related to the phenotypic response. The OPLS/OPLS-DA models provide a ranking of variable influence of projection (VIP) measures, that express the association of the features with the predictive components. Several studies set VIP>1 as a threshold to consider a feature as potential biomarker. Support vector machines (SVM) is a supervised machine learning algorithm appropriate for the interpretation of non-linear data. In contrary to PCA and PLS models, SVM recognises non-linear relationships and identifies vectors or samples in the margin between two groups in order to detect the maximum margin hyperplane [42].

The features of higher statistical impact provided by MVA models, are usually further verified, regarding their statistical importance, with Receiver Operator Curve (ROC) analysis. ROC analysis is used to evaluate the accuracy of a diagnostic biomarker, i.e., to quantify how accurately a metabolite can discriminate to phenotype states. The ROC analyses the trade-off between the true positive rate (TPR) against the false positive rate (FTR) across varying cutoffs of positivity. The area under the curve (AUC) is the interpreted result of ROC curve analysis and represents the sensitivity and the specificity of the diagnostic feature[43].

As the MVA models are utilized for group classification/ prediction and variable selection, they must be thoroughly evaluated regarding their metrics of performance. This is accomplished using the figures of merit, that represent objective measures. I.e., R2 and Q2 are values that assess the fitting ability and the prediction power of models as PLS-DA.

There are several tests commonly used for purposes of model validation i.e., test of permutations, CV-ANOVA and also there are parameters as Hoteling's T2, a multivariate generalized t-test, that are used to evaluate the dataset and point out outliers.

The test of permutations estimates the risk of a PLS-DA (or PLS) models to be overfitted, meaning that the model fits the training set well but, is not able to predict the test set. In this method, the R2 and Q2 provided by the original model are compared to those provided by new models that are generated by random permutation of Y-observations while the X-matrix remains unchanged.

2.7.2.2. Innovative methods for variable selection

The measures resulted from multivariate/univariate models as PLS-DA (VIP), PLS (regression coefficient, RC), t-statistics (p.value), ROC curve analysis (AUC), etc., are commonly used to rank the metabolites (variables) by their statistical importance and are subjected to further investigation, i.e., identification. Although, it is deemed essential to enrich the common-applied biomarker selection strategies, taking into consideration the following issues. At first, the selection of a set of variables based on an optimal performance, e.g., VIP>1.5, underestimated the existence of interactions between the variables.

VIP scores for example reflect the importance of a specific variable, when all variables are included in the model. Thus, these classification models are benefited by the inclusion of variables' interactions during their development, but the final distinction of some variables, may be biased, as it is not ensured that they will show consistent performance for another variables set [44]. An important type of interaction may be the synergetic or joint effect, which is the cooperative performance of a subset of variables. The above-described manner of variable selection efforts to determine a minimal set of the strongest predictors that are linked to the phenotype of interest, also known as minimal-optimal problem. The complementary approach is the detection of all the variables that are relevant to the scope (all-relevant methodologies), including even the weak and the redundant variables, but avoiding the noisy and irrelevant information. All-relevant methods facilitate the comprehensive understanding of the phenotype of interest, and are promising as they can detect dysregulated metabolic pathways, showing the impact of the phenotype in all participating metabolites [45].

It is referred that many machine learning algorithms show decreased accuracy when the number of the input data is far higher than the *optimal* number. Therefore, the introduction of variable selection steps before the classification model, is a new trend in metabolomics, that minimizes the data, increase the accuracy and also speeds the algorithms performance[46].

2.7. KODAMA

Knowledge discovery by accuracy maximization, KODAMA, is a machine learning algorithm based on the principles of dimensional reduction. KODAMA is considered as variable extraction, and not variable selection, methodology, extracting information from noisy and highly dimensional data. The unique characteristic of this algorithm is that ensures the reliability of the given results by conducting a validation procedure of the classification models in the method itself. The core and novel idea behind KOMADA is that a classification model, i.e., a predictor based on clustering, can be refined with the rejection of samples that are not correctly predicted in the cross validation. KODAMA generates an initial clustering and then attempts to refine it through an iterative procedure to increase the cross-validated accuracy. In order do so the algorithm changes the

class labels of the non-correctly predicted samples with their predicted class value. The initial clustering is generated by classification methods, e.g., PCA, or considering that each sample belongs to a different class. KODAMA *considers* the potential occurrence of more than one possible classifications inside the dataset, that show high cross-validation accuracy. To address this fact, the detection of the best classification is performed via a Monte Carlo (MC) procedure, continuously remodeling the classification itself. The cross-validated accuracy is estimated with kNN or PLS methods. The KODAMA pipeline results in suboptimal solutions and is repeated to average the effects provided by randomness [47],[48].

KODAMA is a powerful tool in the investigation of unknown classification trends; however, it can be driven by external supervised information as well, operating as a semi-supervised procedure. Semi-supervised KODAMA performs in two ways. In the first case, the external information is provided partially for only some samples, e.g., disease status, and is used to achieve the maximization of the cross-validated accuracy, without changing the class of these samples. In the second case, the algorithm faces the samples as organized groups, i.e., existence of replicates. So, it links the same samples in a manner that if one of them changes during the cross-validation procedure, the linked ones must change too. In other words, the same samples are forced to belong to the same class [47],[48]. The steps of the KODAMA algorithm are summarized in Figure 9.

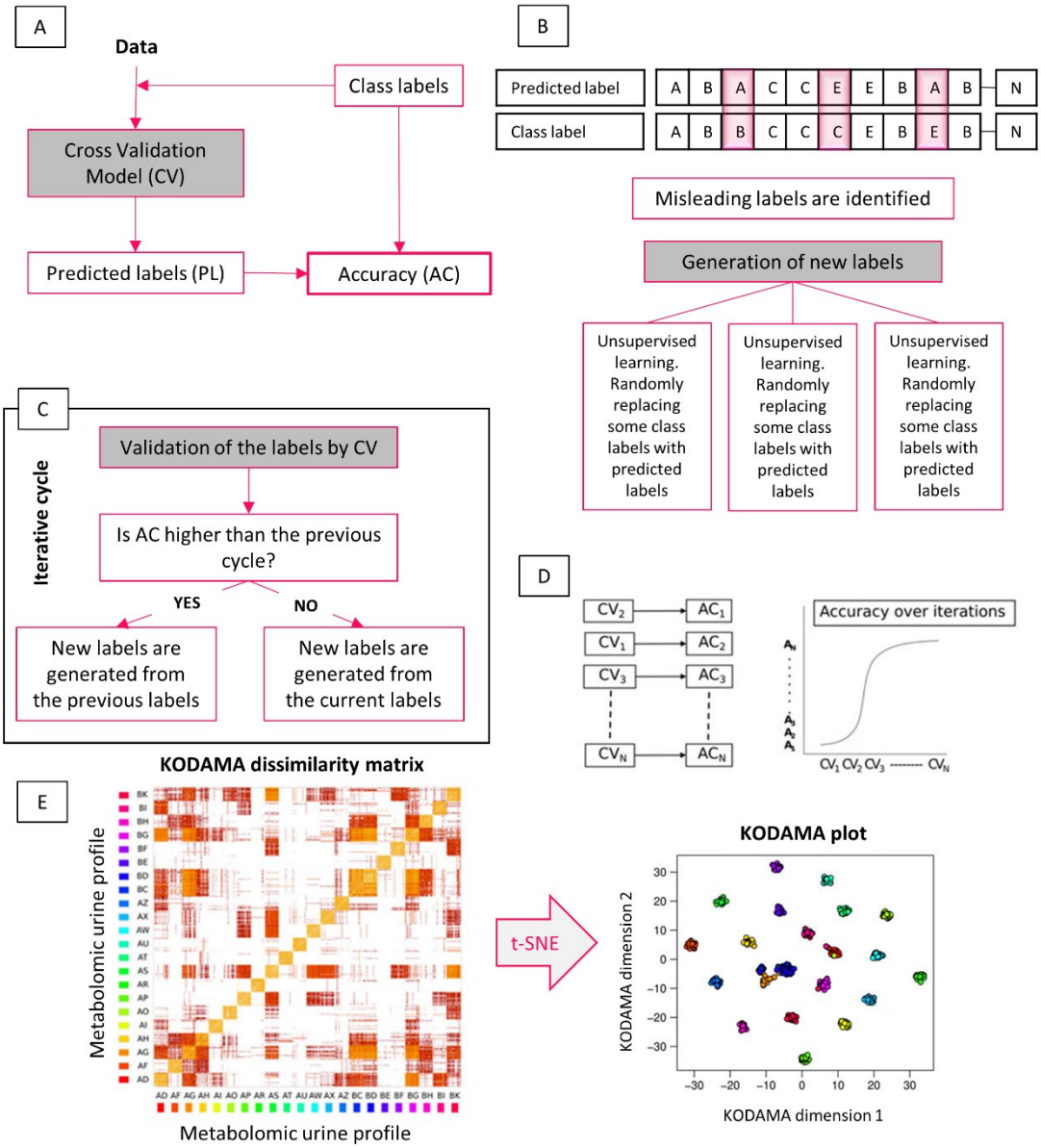

Figure 9 Representation of the steps followed by KODAMA routine (A) Cross-validation model (CV) generates predicted labels (PLs) that are used to calculate the accuracy value (AC). (B) Generation of new labels to conduct the process of accuracy maximization can be i) an unsupervised method, randomly swapping some class labels of misleading samples with predicted labels; ii) semi-supervised type-I, changing only predefined labels and maintaining assigned class labels; or iii) semi-supervised type-II, changing groups of labels together forcing their belonging to the same class. (C) Generation of new labels is an iterative process aimed to identify the labels with the highest cross-validated accuracy. (D) Accuracy values increase with the number of iterations. (E) KODAMA dissimilarity matrix generated as output can be transformed with MDS, or t-SNE, in a low-dimensional space. Figure is modification by Zinga et. al. [48]

2.8. Boruta

The Boruta algorithm belongs to the *all-relevant* variable selection family, attempting to determine all the variables that are in some circumstances relevant to the classification of interest. This approach is promising when the scientific interest focuses on the comprehension of the occurring mechanisms related to the phenotype of interest. The core idea of the *all-relevant* problem is that a variable with weak correlation to the discriminant information does not prove that this variable is not important in conjugation with other variables. The *optimal-minimal* problem used the classification accuracy as the key criterion to identify important variables. However, the same approach cannot be used for the *all-relevant* problem, as the classification accuracy as criterion to reject a variable is unimportant. Wrapper algorithms address the above-described challenge. These are based on specific machine learning algorithms that are trained to fit on a given dataset, Figure 10.

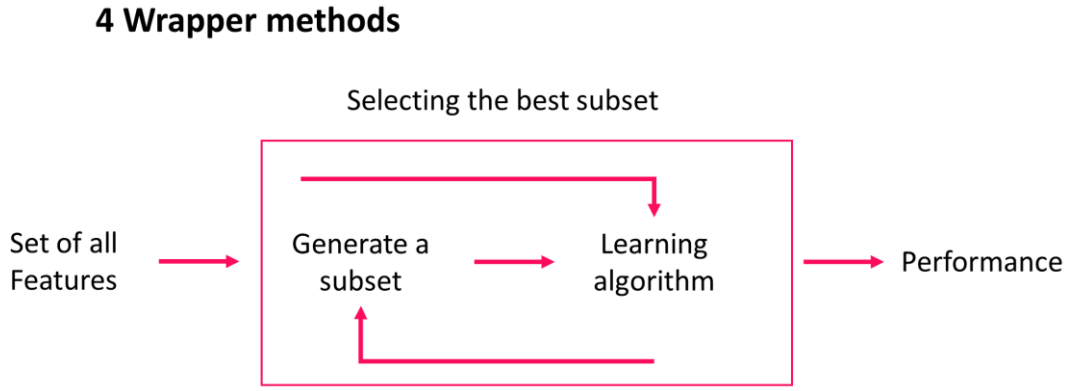

Figure 10 Wrapper variable selection methodology. The figure is modified by "Analytics Vidhya" (https://www.analyticsvidhya.com/blog/2020/10/a-comprehensive-guide-to-feature-selection-using-wrapper-methods-in-python/)

The Boruta algorithm is a wrapper built around the random forest classification algorithm. The key idea of Boruta is that the addition of randomness to the system (input-variables) and collecting the results from the ensemble (total of dummies and true variables) of the randomized samples reduces the impact of random fluctuations and correlations [46],[49].

Boruta includes the following steps:

1. Extend the information system by adding copies of all variables.
2. Rearrange the added attributes to remove their correlations with the response, generating shadow variables.
3. Run a random forest classifier on the extended information system and gather the Z scores computed.
4. Find the maximum Z score among shadow attributes (MZSA), and then assign a hit to every attribute that scored better than MZSA.
5. For each attribute with undetermined importance perform a two-sided test of equality with the MZSA.
6. Consider the attributes which have importance significantly lower than MZSA as 'unimportant' and permanently remove them from the information system.
7. Consider the attributes which have importance significantly higher than MZSA as 'important'.
8. Remove all shadow attributes.
9. Repeat the procedure until the importance is assigned for all the attributes, or the algorithm has reached the previously set limit of the random forest runs [46].

The main loop of the Boruta algorithm is illustrated in Figure 11.

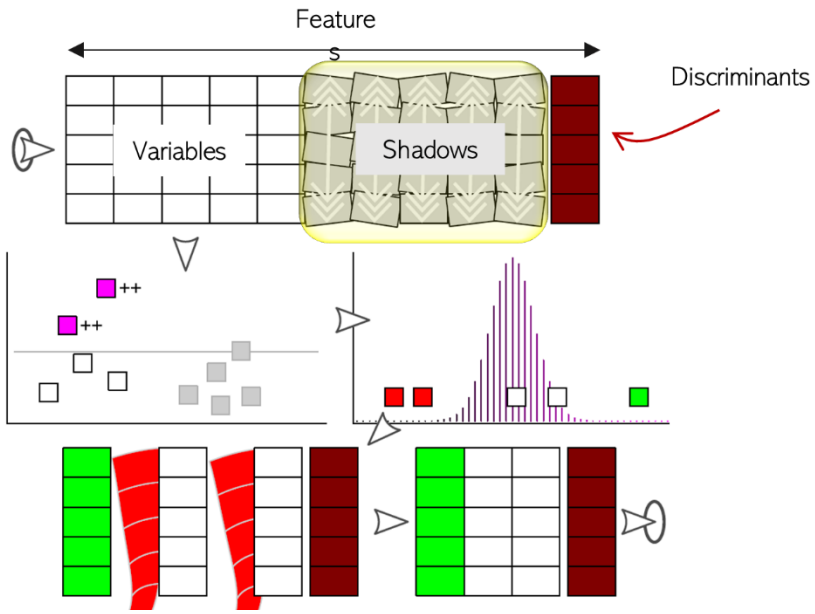

Figure 11 Graphical description of the Boruta algorithm main loop [46].

2.9. MUVR

Multivariate modelling with minimally biased variables selection in R (MUVR) is a variable selection-within-validation algorithm, which performs variable selection within repeated cross validation (rdCV). The algorithm is suggested for datasets that the number of variables exceeds the number of samples (observations), and simultaneously identifies *minimal-optimal* and *all-relevant* variables sets, used for regression, classification, or multilevel analysis. The algorithm attempts to overcome the bias occurring when the variable selection is conducted by using some or all of the samples which were before used for the estimation of the prediction error in the CV process, a fact that is commonly applied to optimize model parameters and to assess model's performance.

The MUVR pipeline is illustrated in Figure 12. During a MUVR loop, the entire data, OUTER segment (variables x samples) are randomly divided into the TEST and the INNER segments. By the end of MUVR loops, all of the samples will be used in the INNER and in the TEST segments. The INNER segment is randomly separated into validation set (VAL) and training set (TRAIN) in order to train the model and optimize its parameters. This step includes recursive ranking and backward elimination of variables. The trained model is then used to predict the outer segment (INNER and TEST), and the procedure is repeated to improve the modelling performance[45].

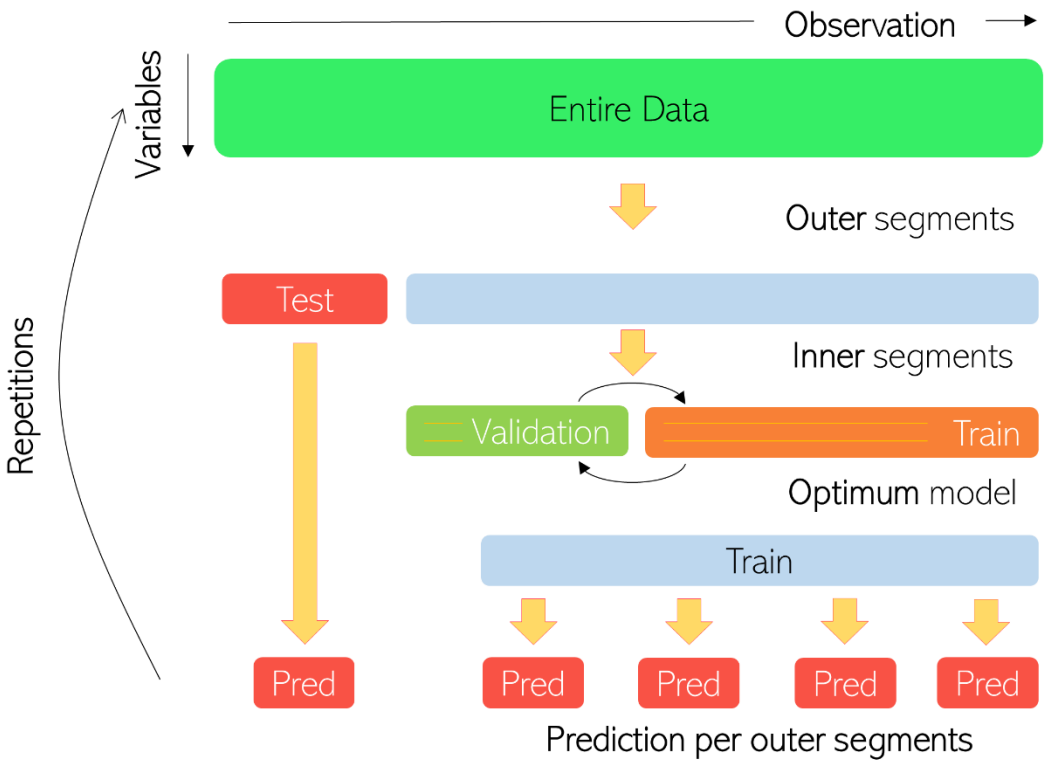

Figure 12 Description of the main steps followed in the MUVR variable selection loop[45].

2.10. VIAVC

In general, random forest (RF) algorithms calculate the variable importance according to the percent increase of misclassification error, when a variable is randomly permuted, and the other variables remain the same. This approach is constrained by the possibility that each variable be given a different probability of sampling. Thus, variable importance analysis based on random variable combination (VIAVC) is another promising methodology that introduced a new variable sampling method, called Binary Matrix Resampling (BMR). BMR ensures that all variables exhibit the same probability to be selected and develops a population of different variable combinations. This method, as the ones mentioned before, considers the synergetic effect among the variables. In VIAVC the importance of each variable is estimated by the percent increase or decrease of AUC (ROC) when the variable is excluded for the modeling. VIAVC classifies the variables in four categories: strongly informative, weakly informative, uninformative, and interfering variables. The loop of VIAVC

identifies all informative variables, rejecting the uninformative and the interfering ones. VIAVC pipeline is consisted by the following steps:

1. Conduction of BMR in the variable space, to create a population of subsets.
2. Use the subsets for sub-models development and assessment of those models performance.
3. Statistical assessment of each variable's importance based on the distribution of variable's inclusion and the distribution of variable's exclusion as well.
4. Determination of the four types of variables (strongly informative, weakly informative, uninformative, and interfering) based on the difference of the two distributions.
5. Removal of the uninformative and interfering variables and retain the informative ones.
6. Repetition of the 1-4 steps until no uninformative and interfering variables exist.
7. Rank the final informative variables based on the P.value of the paired t test on the two distributions.
8. Revealing of the best variable subset with the best prediction accuracy [50].

2.11. 2-way ANOVA, ASCA and MEBA

As mentioned before, metabolomics investigate the composition of the metabolome and the interactions of the metabolites. Except of pointing out discriminating variables, another scientific question concerns the relationships existing between the measured signals. Under this notion, new methods are implemented in the common metabolomics pipelines, aiming to facilitate the interpretation of the experimental observations.

Two-way ANOVA estimates the effect of two independent categorical variables (e.g., time and drug) on a continuous dependent variable (metabolomics signals). In this way, the method determines the main effect of each independent variable and also points out the effect caused by their interaction (if there is one). However, in the majority of the metabolomics investigations

there are more than two known factors that impact the measured variables and therefore a generalization of the above approach is needed.

ASCA (ANOVA simultaneous component analysis) is a generalization of ANOVA aiming to express the metabolomics measurements as a function of more than two effective factors. In other words, when a study investigates the impact of 4 factors, e.g., time, drug, sex, age, in the metabolic profile, ASCA splits up and assigns the observed variation in each individual factors and in their synergetic impact as well.

MEBA (Bayesian time-series analysis) is a method applied for time-resolved metabolomics studies. MEBA estimates the correlations between the observations at different times in order to identify the metabolites that their levels vary the most across each time point.

3. Chapter 3

The Scope

3.1. Scope and Research Objectives

The ever-growing scientific knowledge, the development and exploitation of new techniques and tools and the consolidation of the use of computing resources in research practice, have created promising opportunities for new and more effective strategies in the fields of toxicology and drug safety. This motivating idea played a crucial role in the concept as well as in the progress of the current thesis. The major scope of this research was to contribute to the field of drug toxicity utilizing metabolomics, a novel analytical chemistry methodology. The new trends in toxicology focus on holistic approaches such as the systems biology, to identify the main biochemical causes of drug-toxicity expression and exploit this knowledge in the creation of generalized early prediction models. Under this notion, metabolomics are acknowledged as efficient tools. Particularly HRMS-based metabolomics permit the analysis of the total metabolome, providing the holistic view of the underlying pathology and a comprehensive description of phenotypes of interest. This facilitates the abovementioned pragmatistic approach, as the metabolome is sensitive recipient of internal and external stimuli perturbations. Thus, metabolomics facilitate the comprehension of the biochemistry in toxic and even pre-toxic conditions, pointing out specific biochemical pathways. The determination of specific routs of toxicity, combined with the novel tools of in silico toxicology is deemed as a promising idea for the development of effective and safe drugs. Moreover, metabolomics enable the animal-to-human translation providing enabling the interpretation of the animal based in-vivo testing. Moreover, this thesis focusing on the development and evaluation of thorough metabolomics methodologies, from the peak-picking to the variable selection step, implementing innovative chemometrics methodologies that address particular challenges of metabolomics as the impact of confounders, the determination of latent variables or the DIA identification.

The first objective of the current thesis was the study of metabolomics alterations occurring in toxicity conditions induced by the administration of the

antineoplastic drug Carfilzomib (Cfz). In this study, the in-vivo experiments simulated the Cfz-toxicity state aiming to shed light to the biological causes of toxicity expression. The metabolomics study focused on the determination of toxicity biomarkers, in order to enable the comprehension of the background biochemistry and, to point out the affected metabolic pathways. This knowledge may offer an important contribution in the development of new therapeutic strategies that counterbalance the Cfz adverse effect.

The second objective of the study was the investigation of the toxic impact of the antibiotic colistin. In this case, the in-vivo experiment was planned administered the human doses in laboratory mice aiming to simulate the drug-induced alterations in the non-toxicity state. As colistin is a last resort factor, this study was innovative under the notion that the drug was administered in the real doses (pragmatistic approach), in contrast to other studies that use in-vivo doses that will never be administered to human patients. Thus, in the current case, the metabolic pathways that were altered by the drug, provide a consistent animal-to-human-translation and are relevant to the biochemical alteration occurring in real-time.

The last objective of this thesis was the investigation of the latent alterations occurring in children with malignancies before chemotherapy intervention. This effort aimed to shed light in the unknown personalized factors that combined to the chemotherapy lead to acute cardiotoxicity in children diagnosed with malignancies. This idea was based on the fact that only a percentage of children patients express toxicity. Thus, the determination of metabolomics routes of pre-toxic condition could facilitate the comprehension of the biochemical pathways that, when triggered, lead to cardiotoxicity. This knowledge is considered as crucial in the identification of high and low risk patients and provide opportunities in personalized curation strategies.

4. Chapter

An Untargeted Metabolomics Approach on Carfilzomib-Induced Nephrotoxicity

4.1 Abstract

Background: Carfilzomib (Cfz) is an anti-cancer drug related to cardiorenal adverse events, with cardiovascular and renal complications limiting its clinical use. Despite the important progress concerning the discovery of the underlying causes of Cfz-induced nephrotoxicity, the molecular/biochemical background is still not well clarified. Furthermore, the number of metabo-lomics-based studies concerning Cfz-induced nephrotoxicity is limited. Methods: A metabolomics UPLC–HRMS–DIA methodology was applied to three bio-sample types i.e., plasma, kidney, and urine, obtained from two groups of mice, namely (i) Cfz (8 mg Cfz/ kg) and (ii) Control (0.9% NaCl) (n = 6 per group). Statistical analysis, involving univariate and multivariate tools, was applied for biomarker detection. Furthermore, a sub-study was developed, aiming to estimate metabolites' correlation among bio-samples, and to enlighten potential mechanisms. Results: Cfz mostly affects the kidneys and urine metabolome. Fifty-four statistically important metabolites were discovered, and some of them have already been related to renal diseases. Furthermore, the correlations be-tween bio-samples revealed patterns of metabolome alterations due to Cfz. Conclusions: Cfz causes metabolite retention in kidney and dysregulates (up and down) several metabolites associated with the occurrence of inflammation and oxidative stress.

4.2. Introduction

Carfilzomib (Cfz) is a second-generation proteasome inhibitor, licensed for the curation of relapsed/refectory multiple myeloma. The Cfz drug is distributed to all human tissues except the brain and is rapidly degraded by peptidase cleavage and epoxide hydrolysis, providing non-active metabolites. The metabolism of Cfz is extra-hepatic and independent of the liver function [1]. Despite Cfz's therapeutic activity, the drug has been associated with cardiorenal side effects. However, the interplay and patho-mechanism of Cfz's adverse effects are still elusive. Noteworthily, few preclinical studies have presented potential prophylactic therapies against Cfz's induced cardiorenal

complications, but none of them have been applied into the clinical practice. Therefore, elucidation of the underlying mechanisms of Cfz's cardiorenal effects is an unmet clinical need. Moreover, clinical practice is derived from biomarkers and diagnostic tools that can detect the early onset of Cfz-related cardio-renal complications, and this is an additional need of the hematologic research [1–4].

Concerning nephrotoxicity, several cases of renal failure have been reported, although with no defined hematologic evidence [5]. According to Fotiou et al., the effects of Cfz in kidneys could be described by two potential mechanisms; the first involves the Cfz-induced microvascular toxicity to the renal endothelium, and the second involves the hyperactivation of the complement membrane attack complex [5]. Therefore, an increase in plasma creatinine was reported as an indicator of renal failure.

In an effort to elucidate the mediators of Cfz's nephrotoxicity, **metabolomics** is a powerful weapon. The metabolome is a sensitive recipient of every (internal and external) perturbation, and the study of its alterations is an efficient way to detect and/or describe several pathological conditions, or even their progress. There are two main approaches to metabolome's investigation, as follows: (i) in cases where there is *a priori* knowledge of a condition's metabolic background, targeted metabolomics can determine specific metabolites as potential biomarkers, or (ii) in hypothesis-free cases, the exploratory analysis of the whole metabolome (untargeted metabolomics) could provide significant evidence, enlightening researchers as to the underlying biochemical causes [6–8]. Thus, the untargeted metabolomics approach can reveal hidden aspects of Cfz's related renal dysfunction, as the biochemical mechanism is still non clarified. The existing metabolomics studies on both Cfz nephrotoxicity and cardiotoxicity are still limited. Tantawy et al., have performed the only a multi-omics study, including the plasma untargeted metabolomics as well as focusing on Cfz-induced cardiotoxicity [9].

Our recent targeted metabolomics study was the first report that linked Cfz with renal injury. The results showed extended dysregulation of renal metabolism due to the administration of the drug [10]. Thus, in order to further investigate

Cfz's impact on cardiorenal related metabolism, a complementary untargeted metabolomics approach was employed. The in vivo experimentation was performed by the administration of a translational dose, equivalent to the human therapeutic one, which is previously established to mimic the clinically observed cardiorenal adverse effects in mice [1]. An MS-based methodology was selected to ensure the high sensitivity in metabolite detection. Furthermore, the experimental set up was carefully optimized to the needs of the untargeted metabolomics approach, in order to investigate the "holistic picture" of metabolic alterations. The impact of Cfz was examined in plasma, kidneys, and urine metabolome, aiming to determine how the drug affects the circulatory system. For this reason, the kidneys have been considered as an input/output system between blood and urine excretion. An intriguing aspect of the developed statistical methodology was to reveal relations of the differentially regulated biomarkers that were discovered in the three biomaterials, to uncover inter-circulatory metabolic correlations. Furthermore, attempting to identify as many metabolites as possible, the identification procedure considered potential products of metabolites' metabolism as well, under the notion of the "dark metabolome" [11].

4.3. Results

4.3.1. Data Pre-Processing

Data independent acquisition (DIA) is commonly adopted for untargeted metabolomics, offering both quantitative and structural information by alternately recording low collision energy MS and high collision energy MS (fragmentation spectra) [12]. Despite DIA's advantages, the facile and consistent interpretation of the vast amount of generated data remains challenging. Among a large number of peak picking software, MZmine version 2.51 [13] was selected, as its modular nature provides the ability to evaluate every step of the process while offering the ability to select between a large variety of advanced tools.

The separation of the low and high collision energy (CE) MS spectra was a crucial step for the rest of the procedure, as these two types provide different challenges. The low CE–MS spectra provide information about the intact metabolite structure, while they are also reproducible enough for semi-

quantitation. Low CE–MS information is intended for the biomarkers' determination, so the removal of noise or of non-reproducible peaks is important for the further chemometric analysis, whereas high CE–MS, which are intended to facilitate the identification, are "fuzzier", with lower signal intensity and mass accuracy. Thus, a peak-picking protocol with common settings for both low and high CE–MS would not be efficient. In the peak-picking procedure, the removal of isotopes and adducts, the chromatograms' normalization, the alignment according to retention time (t_R), by employing "well behaved" peaks as standards [14], and the final correction of the aligned list removing the duplicates and gap-filling using the bellow referring algorithms were performed. For the above steps, some parameters were universally set in all datasets and others were adjusted according to the bio-sample type and the ionization mode.

Signal-intensity correction was the penultimate step before conducting the statistical analysis because this study, similar to all long-term metabolomics studies, "suffered" from non-negligible drift of signal intensity attributed to several factors, such as the non-reproducible ESI–LC-MS system, or detector performance [15,16]. To overcome this issue, QC samples were periodically analysed between samples in such a way that three samples were "blocked" by two QC samples. Subsequently, after peak-picking, statTarget2 [17] (utilizing QCs and the QC-RLSC algorithm which employs the LOWESS function), was used to normalize the signal among samples [18]. The results of peak-picking and QC-RLSC signal correction are summarized in Table 2.

Table 2 Summary of the features which resulted from the peak-picking of Cfz HILIC data.

	Number of Variables:		
Dataset	**Detected**	**After QC-RLCS**	**Used for Multivariate Analysis**
Plasma (**+**)	624	346	191
Plasma (**-**)	156	67	49
Kidney (**+**)	1079	964	964
Kidney (**-**)	239	195	195
Urine (**+**)	1769	1509	684
Urine (**-**)	533	458	328

(Detected), those retained from QC-RLSC-based signal correction, performed via statTarget2 (After QC-RLCS), and those included for multivariate analysis that had values different from their median.

4.3.2. Statistical Analysis

A combination of multivariate and univariate analysis was performed for the determination of the variables with the higher contribution to group discrimination. Multivariate analysis was performed using SIMCA 14.1 (Umetrics, Sweden, Upsala). Variables with values which did not differ from the median were excluded before the analysis. The number of features included in multivariate analysis is described at Table 2. Then, PCA analysis was employed to investigate any discrimination trends and possible outliers between the Cfz and Control samples, while PLS-DA models were developed to determine the most differentiative features. Permutation testing was employed to estimate overfitting concerning the Q^2 value. It should be noted that the prediction ability is limited but that is justified due to ethical reasons for the use of lab animals. The results of all datasets' PLS-DA models are summarized in Table 3. The observations that emerged from the multivariate statistical analysis revealed Cfz's influence on all bio-sample types. Referring to kidney-PCA score plots (Figure 13), the two groups are well defined and there is a clear separation between the Cfz and Control samples. In urine, there is a separation tendency, more apparent in the negative mode; however, there is a significant dispersion among samples of the same group. In plasma, the discrimination is quite fuzzy. The PCA score plots suggested that the main effect of Cfz's administration is located in the renal tissue and is partially expressed in urine excretion as well, whereas the impact on plasma is quite vague. The PCA plots from the positive ionization datasets are presented in Figure 13 and the plots of the negative ionization have been attached as Supplementary Figure 1.1.

Table 3 Summary of results from the multivariate and univariate statistical analysis of Cfz HILIC data.

							Number of Biomarkers			
	PLS-DA Model				PLS-DA Permutations Test		PLS-DA Analysis	ROC Analysis	FDR-*t*-Test	Fold-Change
	PCs	Q2	R2	Higher VIP Value	(Q2)	(R2)	VIP > 1.5	AUC > 0.9	*p*-Value < 0.05	log2(FC) > 2
plasma (+)	3	0.5	0.99	2.3	0.32	0.99	11	8	20	107
plasma (-)	4	0.14	0.92	2.53	0.27	0.94	13	0	0	24
kidney (+)	3	0.79	0.99	1.96	0.13	0.98	40	105	110	79
kidney (-)	3	0.93	0.99	1.73	0.38	0.97	21	45	30	29
urine (+)	4	0.86	1	2.35	0.58	0.99	23	68	82	926
urine (-)	3	0.77	0.99	2.49	0.20	0.99	19	41	40	158

The first six columns are referring to PLS-DA classification, namely the number of principle components *(PCs)*, measure of fit *(R2)*, prediction ability *(Q2)*, the higher estimated VIP values *(Higher VIP value)*, and the results of the permutations test *((Q2)* and *(R2)).* The last four columns show the number of variables that exceed the set limit, i.e., 11 variables of the plasma (+) dataset had a VIP value higher than 1.5.

The PLS-DA analysis succeeded in achieving classification between the Cfz group and the Control group for all datasets except plasma (-) and, therefore, the latter was not considered for the feature selection. The kidney showed the highest number of differentiating variables (more VIP values > 1). In an effort to render the analysis more rigorous, only the variables with VIP score higher than 1.5 were considered as potential biomarkers, as they were supposed to exert a higher impact on a group's classification and were submitted for further investigation. A summary of the PLS-DA results, as well as those from the models' validation, are provided in Table 3**.** The PLS-DA score plots of positive ionization are presented in Figure 13**,** while those from the negative are given in the Supplementary Figure 1.2.

Additionally, univariate analysis was implemented, employing MetaboAnalyst 5.0 [19] for ROC curve analysis, an FDR-corrected *t*-test (tt), fold change (FC) analysis, and the creation of volcano plots. The results obtained suggested that the most differentiating variables were detected in the kidneys, which is in accordance with the multivariate analysis. Only the variables with AUC values higher than 0.9 were selected for identification. Therefore, it has been

postulated that the main differentiation due to Cfz's administration is expressed in the kidney metabolome and is also expressed in the urine metabolites. However, neither Cfz nor its administration effects on renal regulation seem to be expressed in blood, as its metabolic composition is essentially not influenced. This may lead to an early conclusion that the effect of the drug is focused on the renal function.

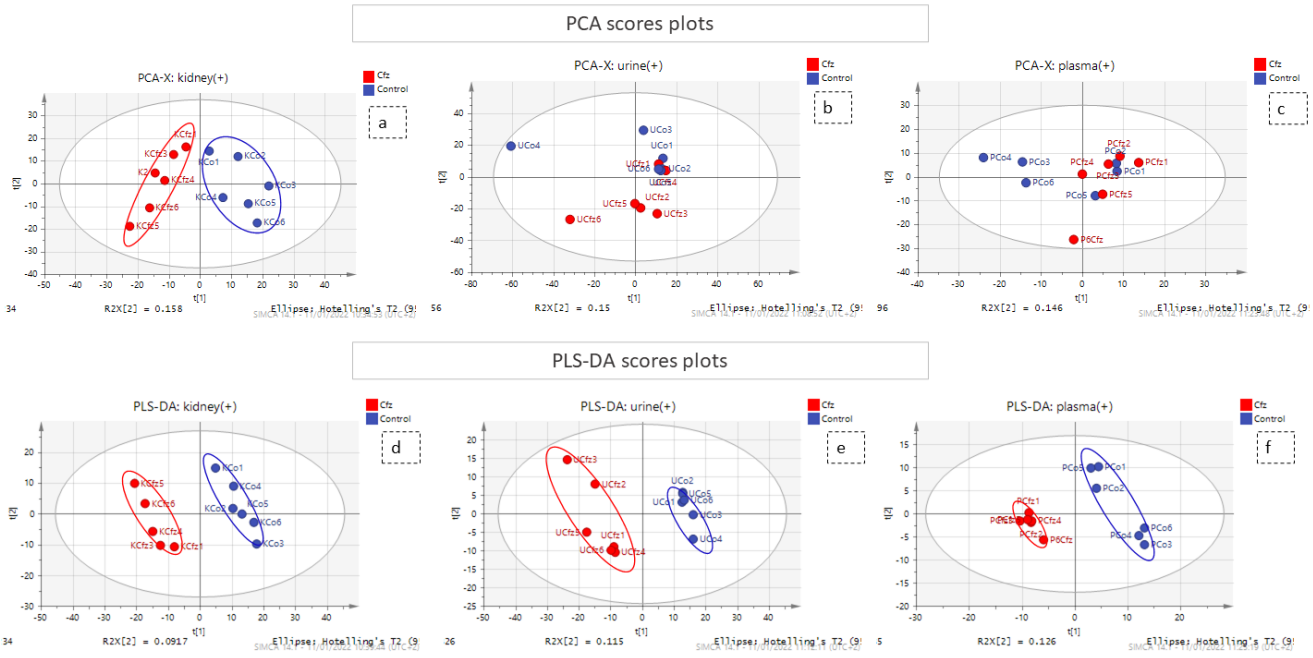

Figure 13 PCA and PLS-DA score plots of Cfz RPLC data. The red and the blue points of the score plots represent the Cfz group samples and the Control group samples, respectively. (a) PCA of kidney (+) dataset; (b) PCA of urine (+) dataset; (c) PCA of plasma (+) dataset; (d) PLS-DA of kidney (+) dataset; (e) PLS-DA of urine (+) dataset; (f) PLS-DA of plasma (+) dataset.

Regarding the steps of the data treatment methodology, the large number of differentiated features resulted from both univariate and multivariate analysis and was limited based on AUC (>0.9) and VIP (>1.5) values; only those were submitted for identification.

4.3.3 Peaks Identification

Concerning DIA, the attribution of peaks to metabolites is a laborious task. As described above, the high CE–MS were used to assign structural features assisting in the unequivocal peaks' identification. However, although this information exists in the high CE–MS, it cannot be accurately related to a

specific precursor ion, since during the bbCID scan mode all precursors are fragmented simultaneously. Thus, the most discriminant features of each dataset (VIP > 1.5, AUC > 0.9) were individually extracted as ion chromatograms (IC) using the DataAnalysis software (Bruker Daltonics, Bremen, Germany) and the background spectra were removed. The cleaned bbCID spectra of the IC were considered as a type of “pseudo-MSMS”, under the notion that high CE–MS info was obtained in a firm retention time range, strictly correlated to the retention time of a low CE–MS spectrum. As many of the discriminant variables could not be identified as metabolites, the experimental pseudo-MSMS was searched alongside the HMDB 5.0 database [20], working towards its implementation in the MyCompoundID (MCID) online library (University of Alberta), http://www.mycompoundid.org/mycompoundid_IsoMS/ (accessed on 20/06/2022). In addition to the human endogenous metabolites, the MCID includes their predicted metabolic products as well. The identification procedure involved 76 features, of which 54 were finally identified. The results concerning plasma (+) showed the lowest number of identified metabolites, and the most differentiating features were considered as formate adducts. Applying the identification workflow in the first four most differentiating features (i.e those with the higher VIP values) of the plasma (+) dataset, it was observed that their mass spectra were identical to those of the calibrant solution, as shown in Figure 14. This is attributed to the endogenous formate which, under the analysis conditions, showed the same MS as the calibrant solution, but was chromatographed as well, showing a t_R at 7.16 min. The total summary of the most differentiating metabolites of the Cfz and Control groups are represented on Table 4.

A total of 71% of the features were annotated, with 75% of them being detected as products of metabolites’ metabolism.

The majority of the identified metabolites are primary or secondary endogenous metabolites, whereas only two of them, *76-perillyl alcohol* and *3-hydroxy-n-methylcarnitine*, belong to nutrients. The last observation may also suggest some differentiation in digestion due to Cfz administration.

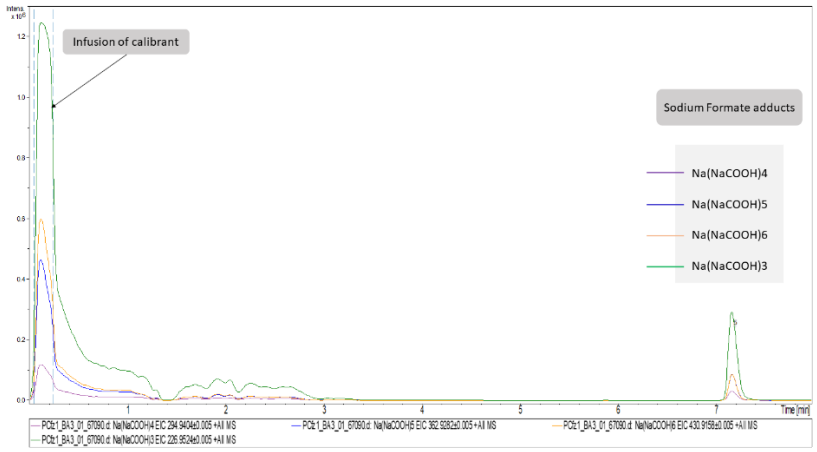

Figure 14 Sodium formate products as important plasma biomarkers in Cfz HILIC data. The first segment of these EICs corresponds to signals obtained from the infusion of calibrant solution. The chromatographic peaks at RT = 7.16 correspond to the adducts of formate detected in Cfz plasma samples.

Table 4 List of identified metabolites, resulted from Cfz HILIC untargeted analysis

Precursor Mass Exp.	Sample Type	ESI Polarity	RT	Cfz-Regulation**	VIP	AUC	Formula	Compound	HMDB	myCID Initial Score	myCID Fit Score	Reactions of Metabolism *	Precursor Type	Precursor Mass Theo.	Error (mDa)
153.0695	urine	+	3.14	↑	1.64	1	$C_7H_8N_2O_2$	N1-Methyl-2-pyridone-5-carboxamide	HMDB0004193	1	0.87	NO REACTION	M + H	153.0659	−0.004
551.2665	urine	+	4.51	↑	1.62	1	UNKNOWN	-							
135.0948	urine	+	6.1	↑	1.47	1	$C_3H_7N_3O_2$	Guanidoacetic acid	HMDB0001528	1	0.52	[+NH_3]	M + H	135.0877	−0.007
286.1103	urine	+	5.78	↑	1.44	1	$C_{11}H_{15}N_3O_6$	N4-Acetylcytidine	HMDB0005923	1	0.72	NO REACTION	M + H	286.1034	0.007
183.1168	urine	+	3.28	↑	1.40	1	$C_9H_{13}NO_2$	p-Synephrine	HMDB0004826	0.96	0.59	[+NH]	M + H	183.1128	−0.004
245.1014	urine	+	3.11	↑	1.40	1	$C_{10}H_{16}N2O_3S$	Biotin	HMDB0000030	1	0.74	NO REACTION	M + H	245.0954	−0.006
181.1626	urine	+	1.35	↑	1.39	1	$C_{10}H_{16}O$	Perillyl alcohol	HMDB0003634	0.99	0.68	[+C_2H_4]	M + H	181.1587	−0.004
214.1852	urine	+	1.36	↑	1.38	1	$C_{13}H_{23}NO_4$	2-Hexenoylcarnitine	HMDB0013161	1	0.92	[-CO_2]	M + H	214.1802	−0.004
199.1739	urine	+	1.35	↑	1.35	1	$C_{10}H_{20}O$	Decanal	HMDB0011623	0.86	0.83	[+C_2H_2O]	M + H	199.1693	−0.005
283.1111	urine	+	5.13	↑	1.32	0.96	$C_8H_{18}N_4O_2$	Asymmetric dimethylarginine	HMDB0001539	0.96	0.55	[+SO_3]	M + H	283.1037	−0.007

112.1154	urine	+	6	↑	1.40	0.96	UNKNOWN	-							
392.2368	urine	+	5.4	↑	1.34	0.9	UNKNOWN	-							
144.9639	urine	-	1.03	↑	1.48	1	$C_6H_4Cl_2O$	2,4-Dichlorophenol	HMDB0004811	1	0.79	[-O]	M-H	144.9606	−0.003
305.1473	urine	-	6.07	↑	1.40	1	$C_{14}H_{18}N_2O_4$	Phenylalanyl-hydroxyproline	HMDB0011176	1	0.844	[+C_2H_4]	M-H	305.1496	0.002
279.0149	urine	-	1.14	↑	1.33	1	UNKNOWN	-							
258.9891	urine	-	4.22	↑			$C_9H_8O_4$	4-Hydroxyphenylpyruvic acid	HMDB0000707	0.99	0.831	[+HPO_3]	M-H	259.0002	0.011
215.0002	urine	-	1.17	↑	1.29	1	$C_8H_8O_2$	Phenylacetic acid	HMDB0000209	1	0.77	[+HPO_3]	M-H	215.0104	0.010
363.0135	urine	-	1.55	↑	1.28	1	$C_3H_6O_3S$	3-Mercaptolactic acid	HMDB0002127	0.81	0.78	[+$C_6H_{11}O_8P$]	M-H	363.0145	0.001
123.0116	urine	-	5.27	↑	1.23	0.9	UNKNOWN	-		.					
324.9654	urine	-	4.75	↑	1.23	0.96	UNKNOWN	-							
365.0294	urine	-	1.12	↑	1.21	0.9	$C_{10}H_{15}N_2O_9P$	Imidazoleacetic acid-ribotide	HMDB0006032	0.86	0.75	[+CO]	M-H	365.0381	0.009
199.9947	urine	-	2.1	↓	1.30	1	$C_3H_8NO_6P$	Phosphoserine	HMDB0000272	1	0.75	[+O]	M-H	199.9955	0.001
230.9946	urine	-	1.37	↓	1.30	1	$C_8H_8O_3$	4-Hydroxy-3-methylbenzoic acid	HMDB0004815	1	0.9	[+SO_3]	M-H	230.9958	0.001
144.0655	urine	-	2.27	↓	1.20	1	$C_6H_{11}NO_2$	Pipecolic acid	HMDB0000070	1	0.76	[+O]	M-H	144.0655	0.000
337.0345	urine	-	1.63	↓	1.29	1	$C_5H_{11}O_8P$	D-Arabinose 5-phosphate	HMDB0011734	0.72	0.8	[+$C_5H_4N_2O$]	M-H	337.0431	0.009
208.9736	kidney	+	8.08	↑	1.38	1	UNKNOWN	-							
349.2322	kidney	+	11.68	↑	1.35	1	$C_{15}H_{29}NO_4$	Octanoylcarnitine	HMDB0000791	1	0.71	[+CO2]	M + NH4	349.2333	0.001
336.1931	kidney	+	6.97	↑	1.31	1	$C_8H_{18}N_4O_2$	Asymmetric dimethylarginine	HMDB0001539	1	0.46	[+$C_5H_3N_5$]	M + H	336.1891	−0.004
133.0617	kidney	+	8.06	↑	1.31	1	$C_3H_8N_2O_2$	2,3-Diaminopropionic acid	HMDB0002006	0.9	0.52	[+CO]	M + H	133.0608	−0.001

245.0777	kidney	+	5.82	↑	1.30	1	$C_9H_{12}N_2O_6$	Uridine	HMDB0000296	1	0.4	NO REACTION	M + H	245.0768	−0.001
160.5245	kidney	+	7.82	↑	1.29	1	UNKNOWN	-							
384.2604	kidney	+	1.59	↑	1.28	1	UNKNOWN	-							
203.1507	kidney	+	11.58	↑	1.28	1	$C_8H_{18}N_4O_2$	Asymmetric dimethylarginine	HMDB0001539	1	0.76	NO REACTION	M + H	203.1503	0.000
297.7149	kidney	+	6.99	↑	1.25	1	$C_{10}H_{20}O_3$	3-Hydroxycapric acid	HMDB0002203	1	0.58	[+$C_5H_4N_2O$]	M + H	297.1809	−0.534
258.4705	kidney	+	7.72	↑	1.24	1	UNKNOWN	-							
166.0867	kidney	+	6.97	↑	1.24	1	$C_9H_{10}O_2$	4-Ethylbenzoic acid	HMDB0002097	1	0.63	[+NH]	M + H	166.0863	0.000
120.42	kidney	+	6.97	↑	1.24	0.97	$C_8H_{11}N$	1-Phenylethylamine	HMDB0002017	0.96	0.71	[-H2]	M + H	120.0808	−0.339
331.1662	kidney	+	6.96	↑	1.23	1	$C_{18}H_{21}NO_4$	(S)-3-Hydroxy-N-methylcoclaurine	HMDB0006921	1	0.7	[+NH]	M + H	331.1652	−0.001
166.8851	kidney	+	6.95	↑	1.23	1	UNKNOWN	-							
171.0176	kidney	+	8.04	↑	1.22	1	$C_4H_8N_2O_3$	Ureidopropionic acid	HMDB0000026	0.77	0.61	NO REACTION	M + K	171.0167	−0.001
600.4706	kidney	+	1.53	↓	1.25	1	$C_{34}H_{68}NO_6P$	CerP(d18:1/16:0)	HMDB0010700	1	0.83	[-H2O]	M + H	600.4751	0.005
556.4439	kidney	+	1.52	↓	1.25	1	UNKNOWN	-							
288.291	kidney	+	6.17	↓	1.23	1	$C_{18}H_{39}NO_2$	Sphinganine	HMDB0000269	1	0.74	[-CH_2]	M + H	288.2897	−0.001
166.4847	kidney	+	6.96	↑	1.28	0.97	UNKNOWN	-							
132.1027	kidney	+	7	↑	1.27	0.97	$C_6H_{12}O_2$	L-alpha-Aminobutyric acid	HMDB0000452	0.98	0.82	[+NH]	M + H	132.1019	−0.001
609.2826	kidney	+	6.15	↑	1.26	0.97	$C_{26}H_{45}NO_8S_2$	Taurolithocholic acid 3-sulfate	HMDB0002580	1	0.59	[+CO]	M + NH4	609.2874	0.005

263.1976	kidney	+	6.94	↑	1.26	0.97	$C_{12}H_{23}NO_4$	Valerylcarnitine	HMDB0013128	1	0.63	[+NH_3]	M + H	263.1965 −0.001
86.3838	kidney	+	6.94	↑	1.26	0.97	UNKNOWN	-						
160.9176	kidney	+	7.82	↑	1.25	0.97	UNKNOWN	-						
179.0616	kidney	-	7.46	↑	1.33	1	$C_6H_6N_4O_2$	1-Methylxanthine	HMDB0010738	1	0.85	[+CH_2]	M-H	179.0564 −0.005
132.0344	kidney	-	7.83	↑	1.26	1	$C_4H_4O_4$	Fumaric acid	HMDB0000134	1	0.74	[+NH_3]	M-H	132.0291 −0.005
225.0678	kidney	-	7.9	↑	1.21	1	$C_7H_{16}NO_2$	4-Trimethylammoniobutanoic acid	HMDB0001161	1	0.73	[+SO_3]	M-H	225.0665 −0.001
130.0914	kidney	-	6.86	↑	1.20	1	$C_6H_{10}O_2$	delta-Hexanolactone	HMDB0000453	1	1	[+NH_3]	M-H	130.0863 −0.005
267.0796	kidney	-	6.6	↑	1.20	1	$C_{10}H_{12}N_4O_4$	Deoxyinosine	HMDB0000071	1	0.92	[+O]	M-H	267.0724 −0.007
124.0114	kidney	-	7.25	↑	1.20	1	UNKNOWN	-					M-H	
180.0716	kidney	-	7.34	↑	1.20	1	$C_9H_8O_3$	Phenylpyruvic acid	HMDB0000205	1	0.82	[+NH_3]	M-H	180.0655 −0.006
289.0737	kidney	-	6.03	↑	1.19	1	$C_6H_{12}O_7$	Galactonic acid	HMDB0000565	0.92	0.74	[+$C_4H_2N_2O$]	M-H	289.0666 −0.007
203.0883	kidney	-	6.78	↑	1.19	1	$C_{11}H_{12}N_2O_2$	L-Tryptophan	HMDB0000929	1	0.85	NO REACTION	M-H	203.0815 −0.007
306.0639	kidney	-	6.03	↑	1.18	1	$C_{14}H_{15}NO_7$	Indoxyl glucuronide	HMDB0010319	1	0.78	[-H_2]	M-H	306.0608 −0.003
296.8881	kidney	-	7.07	↑	1.18	1	UNKNOWN	-					M-H	
171.0116	kidney	-	7.82	↑	1.18	0.97	$C_7H_8O_3S$	p-Cresol sulphate	HMDB0011635	1	0.79	NO REACTION	M-H	171.0110 −0.001
145.0664	kidney	-	7.88	↑	1.16	0.97	$C_5H_{10}N_2O_3$	L-Glutamine	HMDB0000641	1	0.6	NO REACTION	M-H	145.0608 −0.006
303.056	kidney	-	6.61	↑	1.15	1	$C_{10}H_{12}N_2O_8$	Orotidine	HMDB0000788	1	0.93	[+O]	M-H	303.0459 −0.010
128.9636	kidney	-	7.07	↑	1.14	1	UNKNOWN	-					M-H	

164.0767	kidney	-	6.83	↑	1.13	1	$C_9H_{13}NO_3$	Normetanephrine	HMDB0000819	0.87	0.83	[-H_2O]	M-H	164.0706	−0.006
243.0685	kidney	-	6.03	↑	1.10	1	$C_4H_8O_5$	Threonic acid	HMDB0000943	0.96	0.85	[+$C_5H_4N_2O$]	M-H	243.0612	−0.007
379.107	kidney	-	6.07	↑	1.11	1	$C_{15}H_{15}NO_4$	L-Thyronine	HMDB0000667	0.96	0.66	[+$C_2H_5NO_2S$]	M-H	379.0958	−0.011
302.1068	kidney	-	7.68	↑	1.11	0.97	$C_{10}H_{17}N_3O_6$	N2-gamma-Glutamylglutamine	HMDB0011738	1	0.72	[+CO]	M-H	302.0983	−0.009
294.9404	plasma	+	7.16	↑	1.45	1	$Na(NaCOOH)_4$	Formate	HMDB0303296				M+	294.9389	−0.001
362.9282	plasma	+	7.16	↑	1.45	0.97	$Na(NaCOOH)_5$	Formate	HMDB0303297				M+	362.9263	−0.002
430.9158	plasma	+	7.16	↑	1.37	0.91	$Na(NaCOOH)_6$	Formate	HMDB0303298				M+	430.9138	−0.002
226.9524	plasma	+	7.16	↑	1.09	0.91	$Na(NaCOOH)_3$	Formate	HMDB0303299				M+	226.9515	−0.001
332.3335	plasma	+	3.36	↓	1.13	1	UNKNOWN	-							
304.3021	plasma	+	3.44	↓	1.23	0.97	UNKNOWN	-							
326.3804	plasma	+	3.17	↓	1.35	0.94	UNKNOWN	-							
717.0657	plasma	+	6.46	↑	1.36	0.91	UNKNOWN	-							

The identification procedure was performed to those features with VIP and AUC values of > 1.5 and > 0.9, respectively. The MyCompoundID (MCID) online library was used, and both *no-metabolic reaction* and *one-metabolic reaction* have been considered for the identification. Initial and fit score are related to identification efficiency. *Initial score* evaluates the relativity between the theoretical formula and the experimental *m*/*z*, and the *fit score* evaluates the matching between the reference and the experimental or in-silico MSMS spectra. * No reaction refers to metabolites which have not undergone any metabolic reaction, while the entry to the column is attributed to the metabolic reaction i.e., the addition or loss of a corresponding group, ** The (↑) arrow denotes increased levels of the metabolites in Cfz-samples compared to the control and; (↓) arrow denotes the decreased levels of the metabolites in Cfz-samples compared to the control.

4.4 Discussion

During the identification procedure, it was observed that a significant number of features could not be attributed to already known metabolites. Nevertheless, those features did represent real and well-behaved peaks with reproducible signal, mass accuracy, and retention time. Furthermore, they belonged to the most discriminative variables and, thus, their existence could not be ignored as they may belong to the "dark metabolome" [11]. Aiming to annotate more of those features to metabolites, the products of metabolites' one reaction metabolism were also searched as potential renderings through the respective module of the MCID library. This attempt provided associations of the given *m*/*z* (and their MS2 thereof) with metabolites that have been submitted to endogenous enzymatic addition or loss reactions of chemical groups, known to be involved in metabolism.

Several metabolites that are already correlated with cardiovascular diseases and renal disorders were detected as differentially regulated compounds in the kidney and urine samples of the Cfz group.

4.4.1 Asymetric Dimethylarginine

Asymmetric dimethylarginine (ADMA) has been detected in plasma, kidney, and urine samples in three compound forms (*ADMA, ADMA* + SO_3, *ADMA* + $C_5H_4N_2O$). The ADMA is an endogenous metabolite, existing in plasma and tissues, whereas it is appearing in urine as a metabolic product. The compound is produced during protein methylation in the presence of arginine's residues, by protein arginine methyltransferases (PRMTs) [21]. Furthermore, ADMA acts through the inhibition of nitric oxide (NO) production, competing with l-arginine in binding to the active site of nitric oxide synthase (NOS) enzymes [22], and resulting to reduction in NO bioavailability. The NO produced in endothelial cells acts as vasodilator and as anti-atherogenic agent due to its anti-inflammatory and anti-thrombotic activity. The interference of NO synthesis invokes dysregulation of endothelium vascular homeostasis [23], whereas the resulting reduction in NO levels elevates blood pressure and renal vascular resistance. Thus, the increase in ADMA circulating plasma levels is associated with cardiovascular and renal diseases [24]. Therefore, ADMA is considered as a marker of chronic kidney disease or cardiovascular disease. Interestingly, in the

current study, the levels of ADMA in plasma were not statistically different between the two groups but appeared differentiated in the kidney samples, where ADMA was found to be almost two-fold increase in the Cfz group. The increased kidney levels of ADMA imply the increased expression or action of PRMTs or is a sign of elevation of renal metabolic rate. Those observations may describe an instant and probably temporal effect of Cfz on renal tissue, responsible for ADMA's increase and subsequently for the reduction in NO synthesis. This reduction may induce a kind of vascular damage or inflammation and, consequently, acute kidney injury [25].

Mice with ischemia/reperfusion injury showed high renal levels of ADMA [26] and, therefore, the compound may be a marker of this condition. The above study also related ADMA's renal levels with oxidative stress, as the compound was correlated with 8-hydroxy-2′-deoxyguanosine, a marker of oxidative stress.

4.4.2 N1-Methyl-2-pyridone-5-carboxamide

Here, N1-Methyl-2-pyridone-5-carboxamide (2PY) is an end-product of nicotinamide metabolism. The compound has been already associated with uremia and chronic kidney disease (CKD) and is registered as a uremic toxin. The toxicity of 2PY is related with compound's inhibiting activity against poly (ADP-ribose) polymerase-1 (PARP-1) [27]. The PARP-1 participates in several mechanisms, such as differentiation and proliferation, DNA damage repair through chromatin's reshaping, and in the regulation of inflammation, providing cell death or inducing the migration of leukocytes under several conditions. Moreover, PARP-1 is necessary for inducible nitric oxide synthase (iNOS), which promotes NO production [28,29]. Analogously to the ADMA-case, the detection of 2PY implies the reduction in NO production. However, in this case, high levels of 2PY were detected only in the urine of Cfz mice, as the kidney levels of Cfz mice presented a slight increase and their plasma levels were statistically equal. A recent study revealed the correlation between normal renal function and 2PY excretion in urine, namely that the levels of 2PY in patients with kidney damage (renal transplant recipients) were elevated compared to the respective levels of healthy donors. Additionally, the levels of the healthy donors were increased after a kidney donation operation [30]. In the case of

Cfz, the increased levels of 2PY in the urine of treated mice may imply the induction of a locally estimated effect of renal function on the urinary system.

4.4.3 N4-Acetylcytidine

The elevation of N4-acetylcytidine (ac4C) levels in the urine of treated mice was also observed. Here, ac4C is an endogenous nucleoside, a urinary product of RNA catabolism, produced by the action of N-acetyltransferase 10 (NAT10). The urine and blood ac4C levels have been associated with several diseases. According to Jin G. et al., an increased level of ac4C in urine is a sign of inflammatory response. This, combined with the elevation of other modified nucleosides, is observed in patients with uremia. However, ac4C levels in patients with chronic renal failure are decreased. Thus, it is assumed that potential abnormalities of RNA degradation induce irregular accumulation of ac4C in uremic patients. In addition, high levels of ac4C are reported in hypertensive rats. Finally, ac4C increase is associated with oxidative stress in eukaryotes but it is not yet clear if the ac4C increase in urine is the result of this condition [31].

4.4.4 Phenylacetic Acid

Phenylacetic acid (PAA) is a registered uremic toxin, detected in high levels in the urine of Cfz mice. The PAA is a product of phenylalanine's catabolism, increased in the blood of patients with chronic kidney disease and uremia. Furthermore, PAA is an inhibitor of iNOS expression, such as 2PY, and inhibits plasma membrane calcium ATPase. Thus, PAA is suggested to participate in artery reshaping [32]. Additionally, as a uremic toxin, PAA is involved in the activation of polymorphonuclear leucocytes (PMNLs) inducing inflammation. This observation has been verified by in vitro experiments that point out PAA's contribution in the inflammation induction and in the decline of PMNL apoptosis [33].

4.4.5 2-Aminoisobutyric Acid

The 2-Aminoisobutyric acid is an amino acid that has been found to be decreased in the kidneys of treated mice. This compound is reported to prevent kidney tubulopathy, as it inhibits the action of D-serine. The D-isomer of serine is a nephrotoxic agent, causing selective necrosis of the S3 segments of proximal tubules. The D-serine is reabsorbed through the proximal convoluted

tubule of kidney and is degraded, at the same location by the enzyme D-amino acid oxidase (d-AAO), into the corresponding a-keto acid and ammonia. Moreover, the catabolism of D-serine induces the generation of H_2O_2 and is assumed to lead to a decrease in renal cellular glutathione, resulting in high production of ROS and oxidative stress. As the structure of 2-aminoisobutyric acid corresponds to that of D-serine, it is possible that the presence 2-aminoisobutyric acid in the kidney prevents the oxidative stress inhibiting D-serine's catabolism [34,35]. Considering that the decreased levels of 2-aminoisobutiric acid may be involved in kidney injury, provoked by limited inhibition of D-serine catabolism, the correlation between these two compounds in the Cfz and Control kidney mice was studied. The outcome was the detection of increased D-serine levels in Cfz mice, while the compound was not detected in the Control samples, as is represented in Figure 15. The correlation of D-serine and 2-aminoisobutyric acid is also represented in heatmap in the Supplementary Materials. Therefore, the hypothesis that kidney injury caused by oxidative stress conditions induced by the decreased levels of 2-aminoisobutyric acid is verified.

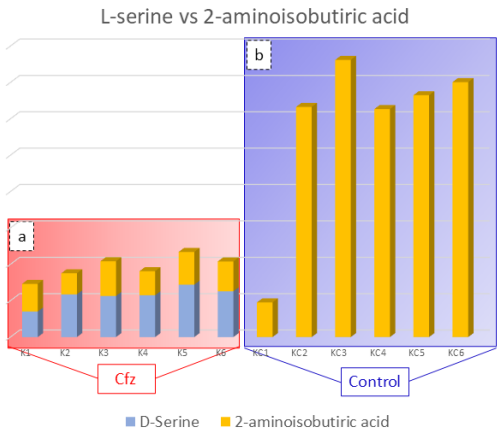

Figure 15 D-Serine and 2-Aminoisobutyric Acid in Cfz HILIC data: Bar charts representing D-serine (blue color) and 2-aminoisobutyric acid (orange color) content in the Cfz (a) and Control (b) kidney samples. Aminoisobutyric acid is decreased in Cfz samples and D-serine is increased, while the later was not detected in Control samples. This suggested that in the Cfz case, the 2-Aminoisobutyric acid is decreased and is not able to inhibit D-Serine from H_2O_2 production and, therefore, cannot protect the kidneys from oxidative stress.

As a general observation, the outcome of the untargeted metabolomics study has revealed several other metabolites that have already been related to renal dysfunction diseases, besides of those described above. These are as follows:

galactonic acid has been detected as a biomarker of CKD [36], threonic acid has been related to oxidative stress induction in patients with membranous nephropathy [37], 2,3-diaminopropionic acid is associated with epithelial cell necrosis of the proximal straight tubules, such as D-serine [38], and octanoylcarnitine, 2-hexenoylcarnitine, and valerylcarnitine, as members of acylcarnitines, are also correlated with AKI [39]. Deoxyinosine is also a nephrotoxicity biomarker [40], whereas indoxyl glucuronide has already been detected in the biofluids of uremic patients [41]. Moreover, some metabolites show differentiation of nutrient's metabolism between two groups which may act on renal function, such as methylxanthine [42].

4.4.6 Exploration of Metabolites Alterations between Different Bio-Samples

As mentioned above, the metabolic profile of kidney and urine datasets seem to be more affected by Cfz administration, in contrast to plasma. In order to explore the potential correlation of metabolites among the kidney as an input/output system, the most discriminant metabolites and their metabolic products were used as "targeted substances" and were semi-quantitated in all the examined bio-samples. The estimated peak areas of the detected metabolites were used to plot the mean metabolite content in every type of bio-sample, along with the standard deviation of the group. According to the observed patterns, the metabolites can be categorized in five potential patterns depending on their variation among plasma, the kidneys, and urine, as follows: (a) metabolites detected in all samples and differentiated only in the kidneys, (b) metabolites detected in all systems and differentiated only in urine, (c) metabolites detected in all systems and differentiated in both the kidneys and urine, (d) metabolites detected and differentiated only in the kidneys, and I metabolites detected and differentiated only in urine. Examples of the different plot patterns are shown in Figure 16. The plasma differences of treated and Control samples were not significant. In addition, the metabolites detected in all samples, were, in the majority of cases, increased in the kidneys of Cfz group. Furthermore, most metabolites are increased in the kidneys and urine of Cfz mice. This observation fosters the idea that Cfz administration increases the renal metabolism of mice. This demands higher consumption of oxygen, a fact that causes hypoxia in parts of the kidney or affects blood pressure regulation

and could be implicated with the manifestation of hypertension. Furthermore, since many of the increased metabolites in the kidneys do not appear proportionally elevated in urine, it is assumed that a significant retention of metabolites occurs in kidneys, perhaps due to alteration of intra-renal metabolite composition, dysfunction in their metabolism, or by water retention in the kidneys [43].

The ADMA showed differences in two types of samples (kidney and urine) and was identified as a feature corresponding to ADMA's mass, and it was also attributed to two features corresponding to products of ADMA's metabolism. Non-metabolized ADMA was detected in all samples and increased in kidneys of the Cfz group, whereas the respective levels of ADMA in plasma and urine Cfz are slightly lower than in the Control. This fact implied that, in addition to ADMA's retention in kidneys, there is also increase in renal biosynthesis in mice treated with Cfz. One metabolic product of ADMA is also increased in the kidneys of treated mice. The compound was excreted as ADMA and ADMA + SO_3, with the latter being detected only in the urine of treated mice, suggesting that the de novo metabolic pathways are triggered in the kidneys under the impact of Cfz The distribution of ADMA in the bio-samples is shown in Figure 17. The total of all ADMA forms in plasma, the kidneys, and urine appears to be 2.4-fold lower in Cfz-treated vs. the Control mice.

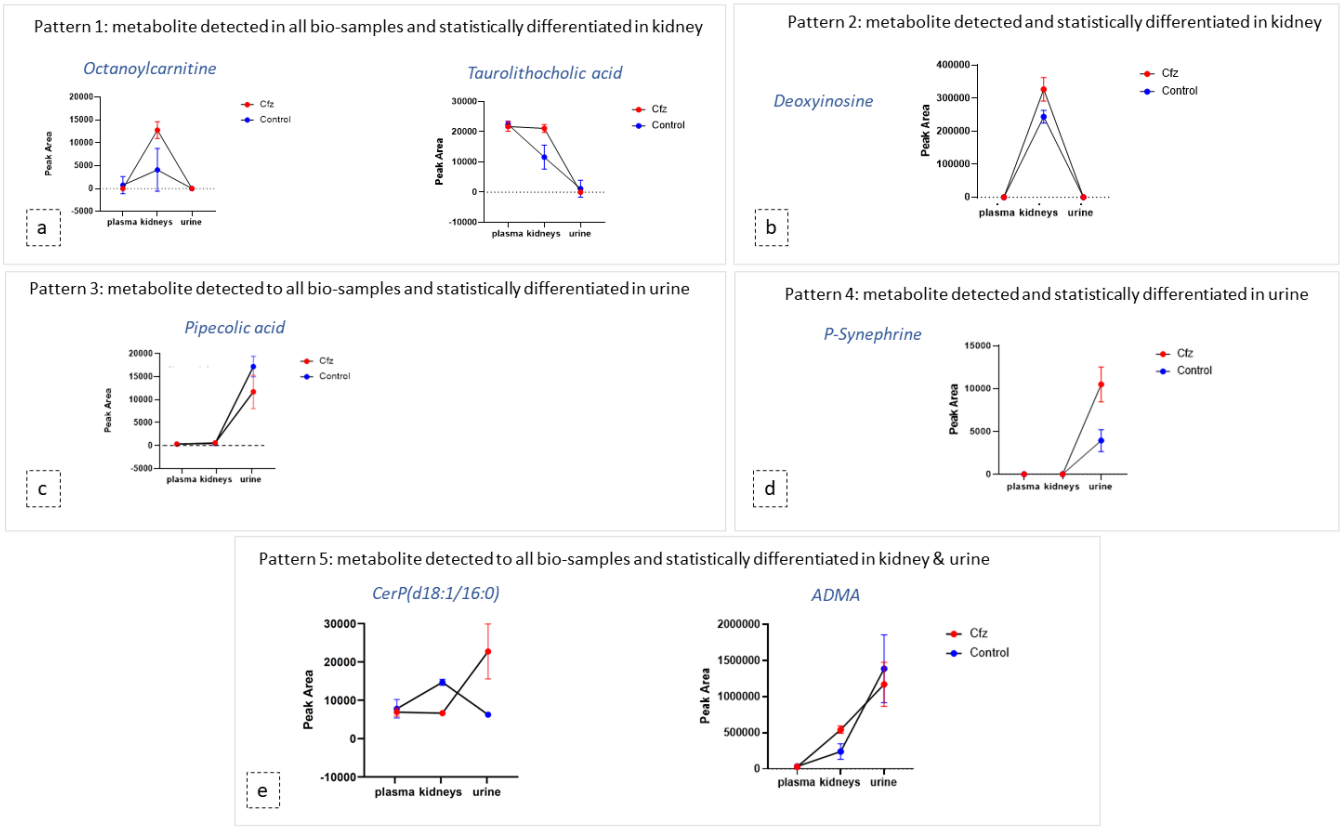

Figure 16 Description of metabolomics interorgan correlation in Cfz HILIC data: Diagrams of metabolite content (expressed via mean and SD) in the plasma, kidneys, and urine of both Cfz (red) and Control (blue) groups. Each diagram represents a type of the revealed patterns of metabolite distribution among bio-samples: (a) metabolites detected in all biosamples and statistically differentiated in kidney; (b) metabolites detected and statistically differentiated in kidney; (c) metabolites detected in all biosamples and statistically differentiated in urine; (d) metabolites detected and statistically differentiated in kidney (e) metabolites detected in all biosamples and statistically differentiated in kidney and urine.

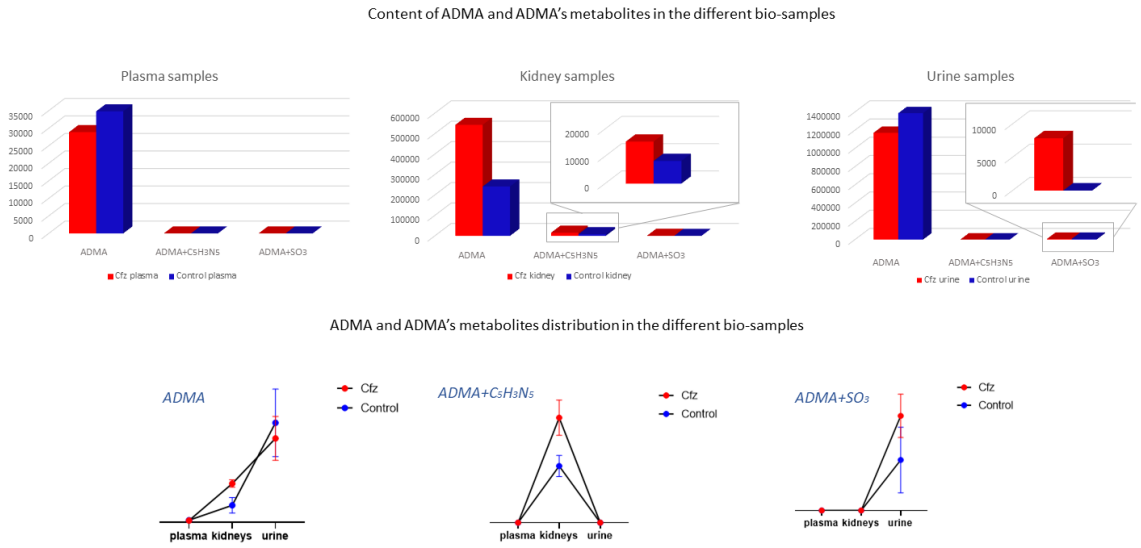

Figure 17 ADMA content in Cfz HILIC data: ADMA had been detected as an intact metabolite and also in two forms of its metabolism. Here, ADMA, per se, is increased in Cfz kidneys; however, it is decreased in Cfz plasma and urine, suggesting an increased rate of ADMA production or ADMA's strong retention at the kidney level. Furthermore, ADMA's metabolites (ADMA + $C_5H_3N_5$ and ADMA + SO_3) have been highly detected in Cfz kidneys and urine.

4.4.7. Discovery of New Potential Biomarkers of Cfz-Related Nephrotoxicity

The current study achieved the goal of determining and identifying more than 40 compounds (metabolites and products of metabolites' metabolism) of proven diagnostic ability (AUC value = 1), that could be potential biomarkers of Cfz-related nephrotoxicity, as follows: kidney (25), urine (16), and plasma (1). So far, only two blood biomarkers (creatinine and urea) are used for the clinical diagnosis of Cfz cardiorenal toxicity [1,44]. Thus, the discovery of this number of potential biomarkers may have a great impact in the prediction of Cfz's renal adverse effects and, therefore, those compounds should be verified in clinical samples.

4.5 Materials and Methods

4.5.1 Sample Collection and Storage

This study employed plasma, kidney, and urine samples of 12 male C57Bl/6J (13–14 weeks of age) mice. The laboratory animals were bred and housed in the Animal Facility of the Biomedical Research Foundation, Academy of Athens. All in vivo experiments were carried out in accordance with the "Guide for the care and use of Laboratory animals" and experiments were approved by the

Ethics Committee (Approval No: 182464;14-05-2019). The mice were housed and maintained according to the ARRIVE guidelines [45]. The animals were randomized in two groups (n = 6 for each group) as follows: i) Control (NaCl 0.9%), ii) Cfz (8 mg/kg) for 6 days [1]. The NaCl and Cfz were injected intraperitoneally on alternate days, and at the end of the experiments mice were euthanized by a high dose of ketamine (100 mg/kg) and subsequent cervical dislocation. Mice were placed in metabolic cages for 24 h for urine collection, and they were provided with food and water ad libitum. Plasma samples were collected by centrifugation of heparinized whole blood at 5000 RPM for 15 min. The bio-samples (plasma, kidneys, urine) were collected at the end of the experiments and stored at −80 °C. Carfilzomib regimens were based on our previous study addressing its cardiotoxicity and are translationally equivalent to human doses [3]. Briefly, in humans, Cfz initial dosing is selected to be 27 or 56 mg/m^2 and can be reduced to 15 mg/m^2 upon manifestation of life-threatening cardiorenal adverse events, before discontinuation of the therapy. In a translational scope, the dose regimen selected for the four-dose protocol is equivalent to a HED of 29.65 mg/m2, which is within the range of the initiation dose of Carfilzomib.

4.5.2 Reagents and Solutions

All the reagents used were of high purity. Methanol and acetonitrile (LC–MS grade) were purchased from Merck (Darmstadt, Germany), ammonium formate was from Fischer Scientific (Geel, Belgium), and formic acid was from Sigma-Aldrich (Steinheim, Germany). Yohimbine hydrochloride and reserpine pharmaceutical grade and primary standards were used as internal standards and were purchased from Merck (Darmstadt, Germany). Distilled water was produced by a Milli-Q purification apparatus (Millipore Direct-Q UV, Bedford, MA, USA).

The preparation of the mobile phase was as follows. For the positive ionization mode, the mobile phase A-pos was an aqueous solution of 5 mM ammonium formate, acidified with 0.01% formic acid, while the mobile phase B-pos was a buffer consisting of acetonitrile–water (95:5 *v*/*v*) containing 5 mM ammonium formate and acidified with 0.01% formic acid. For the negative mode, the mobile phase A-neg was an aqueous solution of 10 mM ammonium formate, and the

mobile phase B-neg was a buffer consisting of acetonitrile–water (95:5 *v*/*v*) containing 10 mM ammonium formate.

The preparation of the internal standard (IS) mix solution was as follows. Two stock solutions for yohimbine hydrochloride and reserpine were prepared using ultra-pure water, with a final concentration of 10 mg/L. Both stock solutions were used to prepare the final mixed IS standard solution, consisting of acetonitrile–water (95:5 *v*/*v*) with 1 mg/L final concentration of yohimbine and reserpine.

For the instrument calibration, a calibrant solution of sodium formate dissolved in 2-propanol water (1:1 *v*/*v*) was employed.

4.5.3. Sample Preparation

Samples from three bio-samples, i.e., plasma, kidney, and urine were employed. Different experimental protocols were implemented for the extraction of metabolites from each sample type. In order to avoid the discrimination of some metabolite classes, the sample pre-treatment protocol involved only a protein precipitation step.

The plasma extraction procedure was as follows: 600 µL of frozen methanol was added to 200 µL of sample and mixed by vortexing for 20 s, before being centrifuged using a NEYA 16R centrifugation apparatus (REMI, Mumbai, India) at 10,000× *g* rpm, 5 min, 4 °C. A 350 µL aliquot of the supernatant was evaporated to dryness by a HyperVAC-LITE centrifugal vacuum concentration (Hanil Scientific Inc, Gimpo, Korea). Samples were stored at −80 °C, and reconstituted before the analysis with 150 µL of IS mix solution [16,18,46,47].

The urine extraction procedure was as follows: 500 µL of the sample was centrifuged (10,000× *g* rpm, 5 min, 4 °C) to precipitate particles. The supernatant was diluted with 1000 µL of a methanol–water solution (1:1 *v*/*v*) and an aliquot of 600 µL was evaporated to dryness, stored at −80 °C, and reconstituted with 150 µL of IS mix solution [18,48–50]. The acquired data were corrected using the total volume of excreted urine of each mouse.

The kidney extraction procedure was as follows. Kidneys were weighted and mixed with an appropriate volume of a methanol–water solution (1:1 *v*/*v*), adjusted to the sample's weight; for every 100 mg of tissue 1000 µL of solution

were added. The sample was homogenized using the tissue homogenizing CKMix lysing kit (Bertin Corp., Rockville, MD, USA) and the CRYOLYS EVOLUTION tissue homogenizer (Bertin Instruments, Rockville, MD, USA). Homogenization was accomplished in two rounds; initially the sample tissue with the half of the aforementioned solution was submitted to the “hard” mode (9600× *g* rpm, three 20 s cycles followed by 60 s pause) of the homogenizer, and then the blend was centrifuged at 10,000× *g* rpm for 10 min and the supernatant was placed in a 10 mL falcon. The rest of the solution was added in the homogenizing tube with the tissue remainder and submitted to a second cycle of a “soft” mode (5000× *g* rpm, one 60 s cycle) homogenization. After centrifugation, the supernatant was mixed with the one obtained by the first homogenization cycle and vortexed for 10 s. An aliquot of 500 µL of the total extract was evaporated until dryness, stored at −80 °C, and reconstituted with 150 µL of IS mix solution before the LC–MS analysis.

4.5.4 UPLC-ESI-QTOFMS Analysis

The chromatographic separation was accomplished with an ACQUITY UPLC BEH Amide column, 2.1 × 100 mm, 1.7 µm (Waters, Ireland, Dublin), equipped with an ACQUITY UPLC BEH Amide VanGuard Pre-column, 1.7 µm, 2.1 mm × 5 mm (Waters, Ireland, Dublin). The data were acquired by implementing the Dionex UltiMate 3000 RSLC UHPLC system (Thermo Fischer Scientific, Dreieich, Germany) coupled to a Maxis Impact QTOF mass spectrometer (Bruker Daltonics, Bremen, Germany) through an electrospray ionization source (ESI) capable of both positive and negative ionization. The column temperature was maintained at 30 °C. The gradient elution program is the same in both ionization modes. The conditions of liquid chromatography and the settings of ESI–QTOF instrumentation are described in Table 5**.**

Table 5 UPLC, ESI and MS conditions for Cfz HILIC analysis

UPLC Gradient Conditions		
% B-(pos/neg)	**Time (min)**	**Flow (mL/min)**
100	0	0.2
100	2	0.2
5	10	0.2
5	15	0.2
100	15.1	0.2
100	22	0.2
ESI–QTOF Settings		
Nebulizer gas		N_2
Nebulizer gas pressure		2 bar
Drying gas		N_2
Drying gas flow		10 L/min
Drying temperature		200 °C
Capillary voltage—positive		3.5 kV
Capillary voltage—positive		2.5 kV
End plate offset		0.5 kV
bbCID collision energy		24 V–36 V (ramp)
***m*/*z* scan range**		100–1000 Da

4.5.5 Data Acquisition

Before the beginning of data acquisition, the QTOF system was calibrated by direct infusion of sodium formate solution, for the *m*/*z* 100–900 Da range, using the HPC algorithm. The *m*/*z* width was set at 1 mDa and the calibration was acceptable when the score value was higher than 99% and the standard deviation of *m*/*z* error (ppm) was lower than 0.5. For each ionization polarity, all samples were analyzed in the same batch. Three types of QC samples were used, one for each type of bio-sample, made as pooled samples consisting of equal aliquots of all samples from the bio-sample. For each dataset, three QCs were analyzed at the beginning and at the end of the acquisition, while during the acquisition, three samples' injections were followed by one QC injection. The injection volume was 5 μL. The data were acquired by employing the broadband collision-induced dissociation (bbCID) mode, which belongs to DIA methodologies. In the bbCID mode, low/high CE–MS data are recorded in alternating scans.

4.5.6 Data Pre-Processing

Six complete studies were performed i.e., three types of bio-samples in positive and negative ionization, namely plasma (+), plasma (-), kidney (+), kidney (-), urine (+), urine (-), thus, six datasets were submitted to further processing.

Data collected from DIA were processed as shown in the Figure 6. Samples and QCs were exported as mzXML files using DataAnalysis (Bruker Daltonics, Bremen, Germany) and imported into MZmine 2.51 [13]. Initially, the information corresponding to high collision energy were cropped and removed from the processing file. The data treatment procedure was applied to the low collision energy spectra, as this piece of information represents the biomarkers in their intact form. The peaks corresponding to isotopes and adducts were excluded from further statistical analysis but retained for the annotation. The MZmine parameters are summarized in Table 6.

Table 6 algorithms and settings applied for the untargeted peak-picking of Cfz HILIC data.

Step	Algorithm	Parameters	Comments
Mass Detection	Centroid Mass detector	Noise level: 500–1500	adjusted to each dataset
Chromatogram building	ADAP Chromatogram builder	Number of scans: 10	
		Group Intensity threshold: 1000–2500	adjusted to each dataset
		Minimum highest intensity: 1000–4000	adjusted to each dataset
		m/*z* tolerance: 5 mDa	
Chromatogram deconvolution	Noise amplitude/AUTO mz centre calculation	Minimum peak height: 1100–4100 abs	adjusted to each dataset
		Peak duration range: 0.1–1 min	
		Amplitude of noise: 100–1000	adjusted to each dataset
Isotopes	Isotopic peaks grouper	*m*/*z* tolerance: 5 mDa	
		Retention time tolerance: 0.1	
		Maximum charge: 2	
Adducts	Adducts search	[M + Na], [M + K], [M + MeOH], [M + HCOOH], [M + ACN], etc.	
		RT tolerance: 0.1 min	
		m/*z* tolerance: 10 mDa	
		Max relative-adduct peak height: 100%	
Normalization	Retention time calibration	*m*/*z*: tolerance 5 mDa	
		Retention time tolerance: 0.2 min	
		Minimum standard intensity: 5000–10,000 abs	adjusted to each dataset
Alignment	Join aligner	*m*/*z* tolerance: 10 mDa	
		Retention time tolerance: 0.5 min	
Remove duplicates	Duplicate peak finder/New Average	*m*/*z* tolerance: 20 mDa	
		RT tolerance: 0.8 min	
Gap filling	Peak Finder	Intensity tolerance: 20%	
		m/*z* tolerance: 10 mDa	
		Retention time tolerance: 0.8 min	

The obtained peak lists were submitted to QC-based signal correction using statTarget2 [17]. The QC-RLSC algorithm, based on the locally weighted scatterplot smoothing non-parametric regression (LOWESS), was employed

[18]. Non-zero variables with values lower than 70% were removed from the dataset, whereas the default parameters of QCspan and CV% cutoff were used. The efficiency of signal correction was evaluated by PCA, with all QCs being included to a tight cluster, as shown in Figure 18.

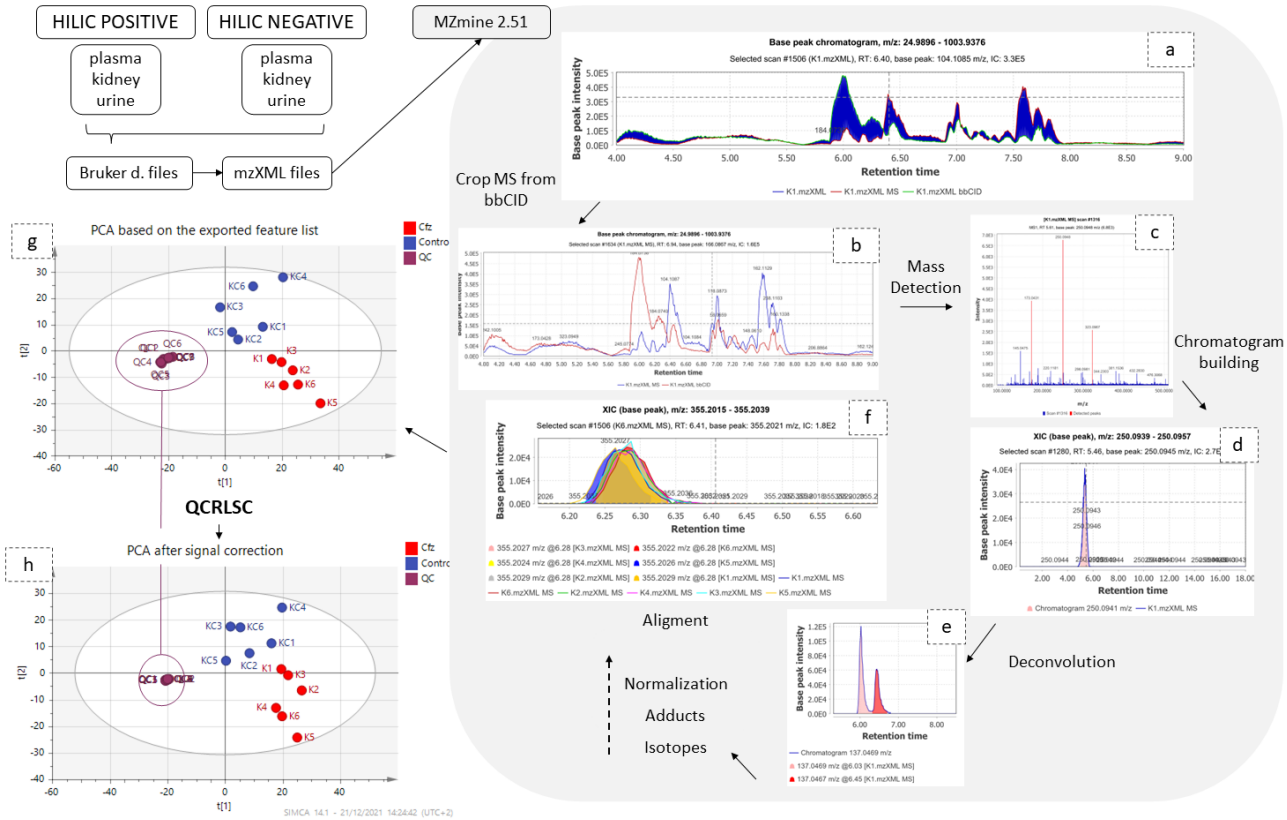

Figure 18 Graphical description of data pre-processing workflow. The low and high CE–MS before (a) and after (b) their separation; (c) set of noise levels for the mass detection; (d) one peak which resulted from the chromatogram building; (e) the chromatogram deconvolution; (f) the result of the alignment; the PCAs before (g) and after (h) signal correction.

4.5.7 Multivariate Analysis

The SIMCA 14.1 software (Umetrics, Sweden) was used for the multivariate analysis. Here, PCA modeling was applied to investigate whether the administration of Cfz affects the metabolic profile of plasma, kidneys, and urine. Subsequently, PLS-DA was employed to point out the discriminant variables (corresponding to mz_t_R features) between the Cfz and the Control groups. For both PCA and PLS-DA, unit variance (UV) and Pareto data scaling, combined with different types of data transformation, were tested. The tested transformation algorithms did not provide any improvement in the normality of

the data and, therefore, no transformation methodology was employed. The UV scaling afforded better clustering; therefore, it was used in all cases. Permutation testing (100 random permutations) was used to evaluate their validity of the PLS-DA models and estimate the degree of overfitting.

4.5.8. Univariate Analysis

MetaboAnalyst 5.0 was used for the univariate analysis which included ROC (receiver operating characteristic) curves, FDR-corrected t-tests, fold change analysis, and volcano plots with fold change threshold [19]. The steps of both multivariate and univariate analysis are shown in Figure 19.

Thank you. We are afraid that we are not able to revised that according to your suggestion, as the software does not provide us access to that settings.

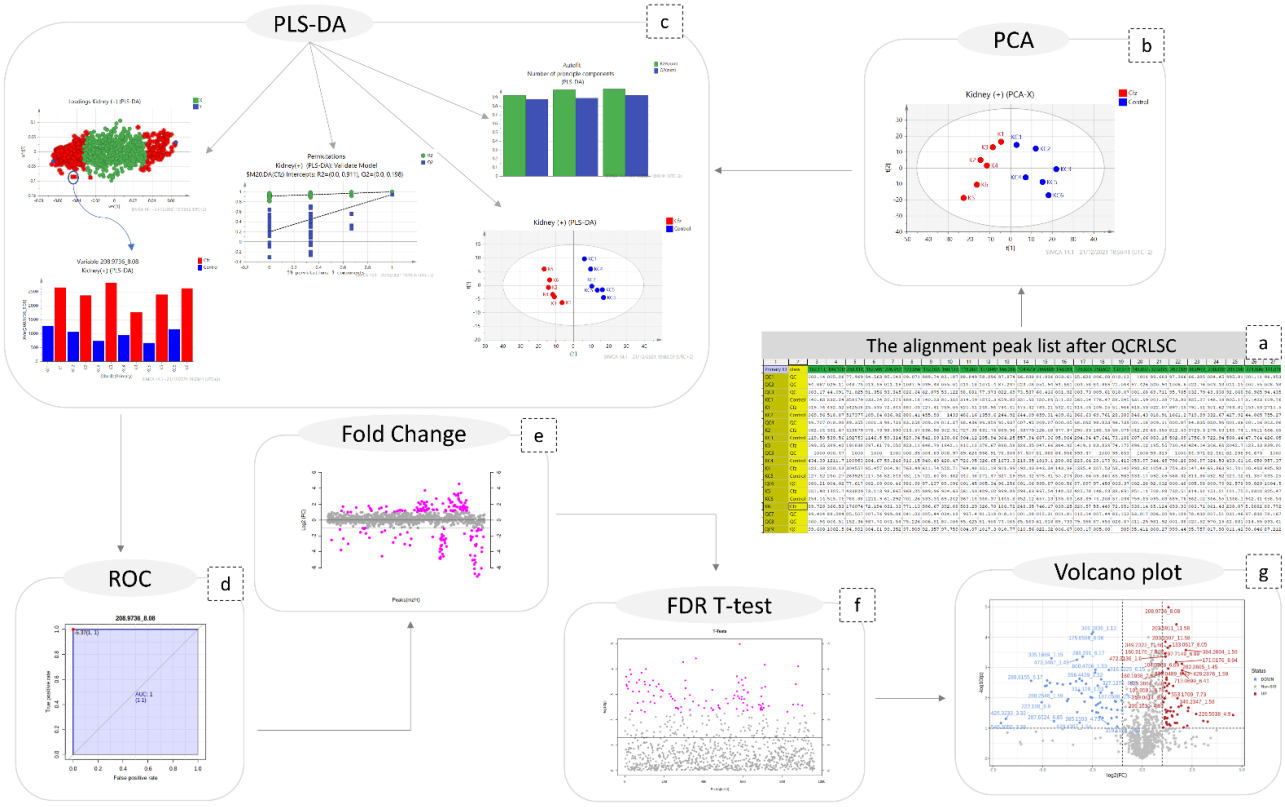

Figure 19 Graphical description of the statistical workflow for kidney (+), as follows: (a) the final peak list; (b) PCA; (c) PLS-DA; (the number of PCs; the scores-plot the loadings plot; the content plot of a discriminant variable; the result of permutations test) ; (d) the ROC curve of a biomarker; (e) the fold change plot; (f) FDR-TT representation; (g) volcano plot.

4.5.9. Peaks Identification Procedure

The most significant variables i.e., those with VIP values higher than 1.5 and AUC values higher than 0.9, were selected for identification. DataAnalysis

(Bruker Daltonics, Bremen, Germany) was employed to extract ion chromatograms with a specific *m*/*z* (tolerance: 10 ppm) and RT (tolerance: 0.2 min) of the selected features and to “clean” their spectra from background noise. The “cleaned” low and high CE spectra of each selected feature were imported into the RamClustR [51] R-based package in order to assign a pseudo-MSMS to their corresponding precursor ions. Each low CE ion combined with its pseudo-MSMS was searched online and confirmed by the comparison to the reference metabolite MS2 spectra, or its in-silico fragmentation, available in MCID and in the HMDB 5.0 database [20].

The myCompoundID online library was used for peak identification, as the library includes information of both human metabolites and of their metabolism-products [52]. All types of precursor ion adducts and both *no reaction* and *1 reaction* mode were investigated. For the experimental pseudo-MSMS spectra, we used the predicted MS2 spectra as provided by the MyCompound ID library, which uses the HMDB and the Evidence-based Metabolome Library (EML). In the cases of the non-metabolized metabolites, the MS2 was confirmed by comparison to the experimental spectra existing in HMDB. As universal requirements, *m*/*z* tolerance was set to 5 mDa for the precursor ions and to 10 mDa for the fragments. The results were evaluated with their initial score (>0.95) corresponding to formula prediction and with their fit score (>0.75), corresponding to a similarity between experimental and reference MS2 spectra. In the cases of tie between two potential metabolites, their biological disposition was considered. The identification procedure is described in Figure 20.

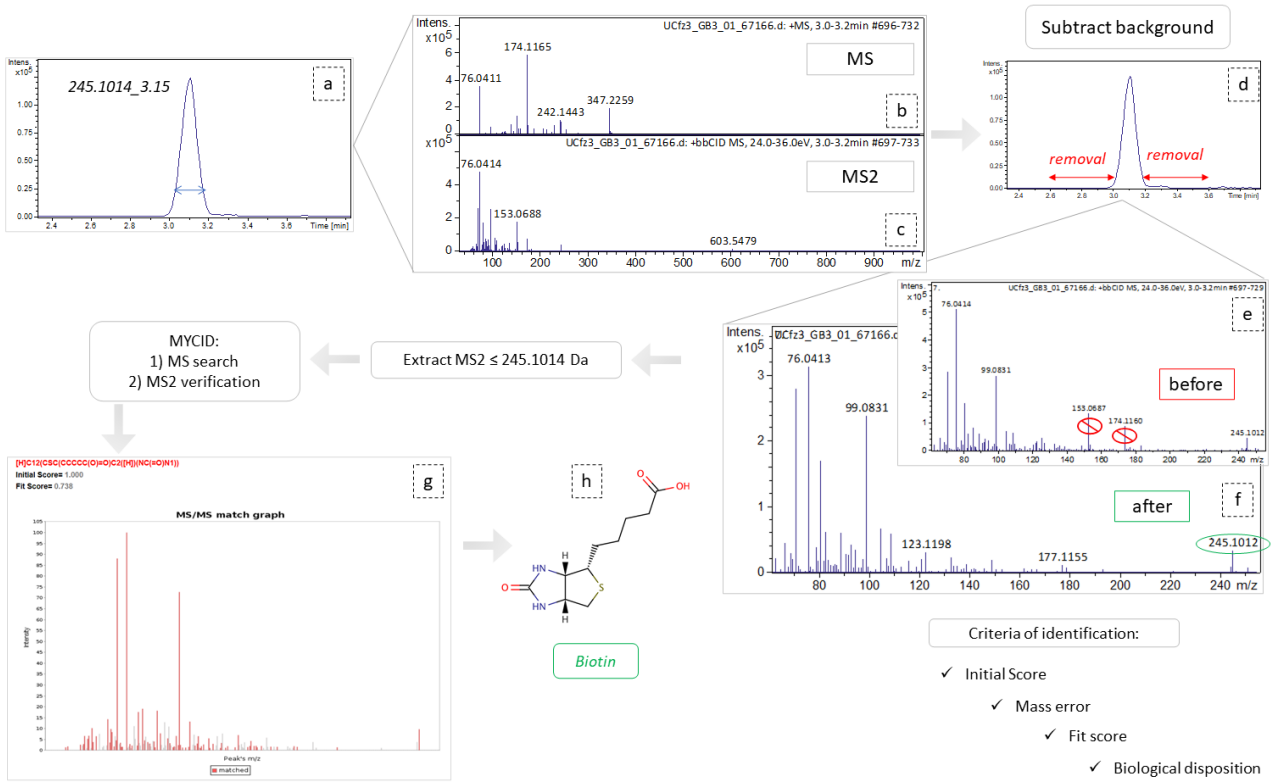

Figure 20 Graphical representation of identification workflow, employing the example of variable 254.1014_3.15. (a) Extraction of ion chromatogram, m/z= 245.1014, tR= 3.14; (b) and (c) refer to MS and MS2 spectra, respectively, corresponding to this peak area; (d) the area removed as background spectra; (e) and (f) represent the MS2 spectra before and after background subtraction, respectively; (g) MSMS match graph obtained from MCID, during the identification procedure; (h) the structural representation of biotin.

4.5.10. Data-Driven Suspect Screening of Metabolites

Inspired from the outcome of the above procedures and attempting to investigate the correlations of metabolites among the plasma, kidneys, and urinary system, a hypothesis driven metabolites determination was designed. One list of “suspect-compounds” was created for each polarity and applied to all samples, regardless the bio-sample. These “suspect-lists” included all the identified metabolites and their possible metabolism products. The TASQ Client 2.1 software (Bruker Daltonics, Bremen, Germany) was used to perform the target screening analysis and the semi-quantitation thereof by peak integration. Each sample was subjected to software-based internal calibration, employing a sodium formate spectrum for the calibration. For the compound detection, mass tolerance was set to 5 mDa, RT tolerance to 0.5 min, and the signal to noise (S/N) level was set at 10.

4.5.11. Exploration of Metabolites Alterations between Different Bio-Samples

The above results were used to plot diagrams with the mean and standard deviation (SD) of the detected metabolites' signal. The mean and the SD were estimated for each group and bio-sample type, aiming to highlight correlations and patterns among the blood, the renal, and the urinary system due to the Cfz administration.

4.6 Conclusions

Carfilzomib is an authorized anti-cancer drug for the treatment of relapsed/refractory multiple myeloma. The drug has been associated with cardiorenal adverse events of unknown pathobiology. Thus, a high-throughput untargeted metabolomics study was implemented to enlighten aspects of nephrotoxicity effects due to Cfz administration. For the metabolomics study, plasma, kidney, and urine samples obtained from treated and Control mice were employed, creating six sets of data (two ionization polarities applied on three bio-sample types).

The results show that kidney and urine were mostly affected, whereas plasma maintained its homeostasis after the drug administration. This fact was confirmed from the statistical analysis, whereas metabolites found up or down regulated were in accordance with an acute renal dysfunction. This was also verified from by the inter-organ correlation. As a general observation, kidney and urine samples of Cfz treated mice show higher numbers of differentially regulated metabolites, indicating a potential increase in the renal metabolism. The inter-organ correlation study pointed out several metabolites, increased only in kidney but not in urine and plasma samples. Furthermore, several metabolites were detected only in the urine samples. Thus, it is assumed that Cfz (i) elevates renal metabolic rate, which may induce high consumption of oxygen, and (ii) provokes retention of metabolites, although this is not expressed in the plasma levels of these compounds.

The identification procedure ended up with 67% of identified features. The identified biomarkers revealed a potential explanation of both Cfz's renal toxicity and the cause of retention in kidneys, as several of them are already associated with renal failure or kidney injury; however, in the literature, the majority of them

referred to blood samples. Thus, it is assumed that Cfz influences renal metabolome in several directions, dysregulating more than one biological pathways, and causing locally estimated damage to renal function. Three main mechanisms seem to co-operate for this local kidney injury, as follows: (a) inhibition of NO production (via ADMA, 2PY, and PAA) resulting to kidney resistance elevation, (b) increase in oxidative stress (through ac4C and 2-aminoisobutyric acid), and (c) inflammation and kidney injury (due to acC4, PAA, 2PY and 2-aminoisobutyric acid) under the influence of the NO decrease. This local damage to kidneys may be the cause of the component's retention.

5. Chapter 5

Metabolomics point out the effects of Carfilzomib on aromatic amino acids biosynthesis and degradation.

5.1. Abstract:

Background: Carfilzomib (Cfz) is an antineoplastic agent indicated for the treatment of multiple myeloma. However, its beneficial action is attenuated by the occurrence of cardiotoxicity and nephrotoxicity as the most common adverse effects. So far, there is well-established knowledge on the pathomechanisms related to these side-effects, but the research on the metabolic alterations provoked by the drug is limited. (2) Methods: An in-vivo simulation of Cfz induced toxicity was developed in i) Cfz-treated and ii) Control mice. An RP-HRMS-based protocol and advanced statistical treatment were used to investigate the impact of Cfz in the non-polar metabolome. (3) Results: The differential analysis classified the Cfz and Control mice and resulted in a significant number of identified biomarkers with AUC > 0.9. The drug impaired the biosynthesis and the degradation of aromatic amino acids (AAA) and led to alterations of uremic toxins in the renal and urine level. Furthermore, the renal degradation of Tryptophan was affected, inducing its degradation via the kynurenine pathway. (4) Conclusions: The renal levels of metabolites showed impaired excretion and degradation of AAAs. Cfz was finally correlated with the biosynthesis of renal dopamine, explaining the biochemical causes of water and ion retention and the increase of the systolic pressure.

5.2. Introduction

Carfilzomib (Cfz) is an antineoplastic agent employed for the cure of the relapse/refractory multiple myeloma, however its action is compromised by high percentages of cardiovascular and renal side effects. Previous studies have shown that Cfz nephrotoxicity is caused by the activation of the mineralocorticoid receptor (MR) through the dysregulation of water/ion transport and the urine electrolyte balance [51]. The existing metabolomics study show that Cfz leads to metabolic dysregulations that are mainly located in the kidneys and urinary system [51],[52]. Besides, the found metabolites prove the correlation between the drug administration with conditions of

inflammation, kidney injury and oxidative stress [52]. However, the knowledge regarding the influence of Cfz on the metabolic pathways of the circulatory and the urinary system remain limited.

Metabolomics are considered as informative tools in drug induced toxicity field of research. Thus, the conjugation of different metabolomics platforms, i.e., reversed phase (RP) and HILIC chromatography, or positive and negative ionization, ensure complementarity regarding the obtained information. The current study, aimed to enrich the existing knowledge employing RP-HRMS conditions to emphasise on the impacts of Cfz in the non-polar metabolome (as those compounds have better chromatographic performance in RP-columns). The final results confirmed the already known conclusions but also showed that the non-poral metabolites exhibit different regulation patters between circulatory and urinary system, in contrary to the polar ones. The identified metabolites showed pronounced impact of Cfz on the renal aromatic amino acids, which are strictly linked to the kidney disfunction. Also, the regulation of renal dopamine was altered providing new evidence on the background of Cfz induced nephrotoxicity.

5.3. Results

The library-based peak-picking workflow (using the in-house DB of metabolites internal standards, enriched with metabolites already linked to Cfz nephrotoxicity) managed to detect 246 and 140 metabolites for ESI+ and ESI- respectively, using lists of 270 and 219 suspect metabolites accordingly. These metabolites were combined with the library-free resulted features (mz_rt) and they were submitted to univariate and multivariate statistical analysis.

PCA analysis, Figure 21 (a,c,e,), described that in all plasma, kidney, and urine datasets the separation trends are in accordance with Cfz administration. However, the data from plasma and kidneys exhibited more pronounced separation, with lower groups inner variance in contrast to urine datasets. Additionally, PLS-DA, Figure 21(b, d, f) confirmed that the developed models succeeded to classify the two groups efficiently. The test of permutations, the misclassification test, and the ROC curve analysis of the models' sensitivity/specificity were used to assess the validity the prediction ability as

well as to check for overfitting. The results of these trials are summarized in Table 7:

Table 7 Summary of OPLS-DA figures of merit for the Cfz RPLC data.

	R2(Y) PLSDA	Q2 PLSDA	R2(Y) PERMUT.	Q2 PEMUT.	AUC ROC' PLSDA	MISS CLAS. ERROR (CFZ) (%) / PLS-DA
Plasma (+)	1	0.93	1	0.89	1	0
Kidney (+)	0.99	0.88	0.98	-0.23	1	0
Urine (+)	0.99	0.94	0.93	0.063	1	0
Plasma (-)	0.98	0.62	0.97	0.32	1	0
Kidney (-)	0.99	0.77	0.93	0.01	1	0
Urine (-)	0.99	0.82	0.93	0.13	1	0

A high number of variables were found differentially regulated, (either up or down), due to the administration of the drug. It is worth mentioning that in the urine case, several features were only detected in Cfz samples, in a reproducible way.

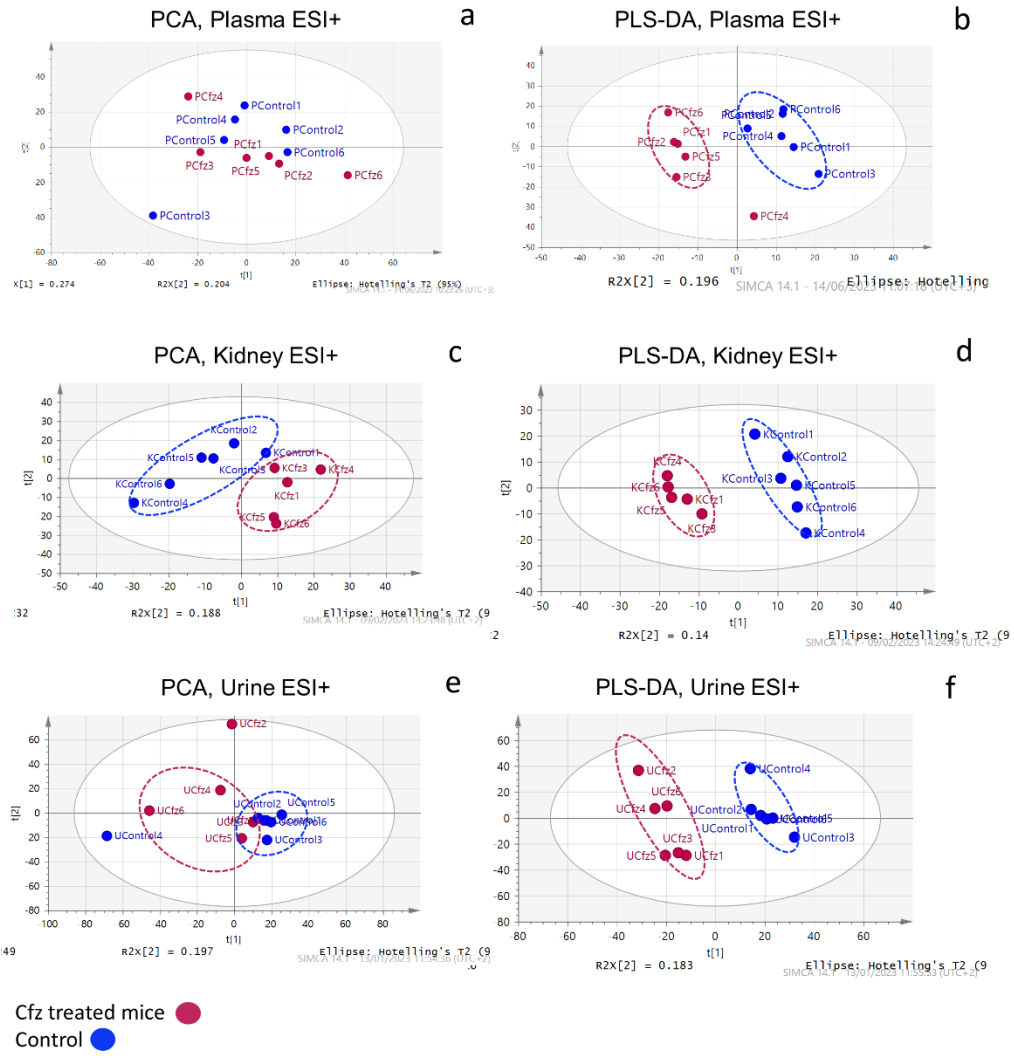

Figure 21 PCA scores plot of plasma ESI + (a); kidney ESI + (c); urine ESI + (e); and PLS-DA scores plot of plasma ESI + (b); kidney ESI + (d); urine ESI + (f).

Afterwards, univariate ROC analysis, applied in raw data, was used to detect potential biomarkers capable to classify a sample as Cfz-induced nephrotoxicity. The AUC values > 0.9 and VIPs >1.5, were used as criteria for variable ranking. The urine dataset provided the higher number of differentiating variables (565), followed by kidneys (191), and plasma (150). The 65% of these important variables were found increased in plasma and kidney of Cfz treated mice, whereas the 65% of them was decreased in the urine of the same mice.

5.3.1. Post-hoc analysis

The increase of metabolites in plasma and kidney of Cfz-treated mice and decrease in the urine of the same mice. This was further investigated, as a potential pattern of metabolites regulation, employing ASCA and MEBA. It is important to note that in this step, the major goal was to investigate the holistic impact of the drug in the circulatory (plasma, kidney) and urinary (kidney, urine) systems, recognizing kidneys as a connection point. The ASCA was used to investigate for important interactions between the drug influence and the biosample. The major patterns resulted by ASCA procedure regarding the differentiation of the variables in the Control and Cfz mice Figure 22a, show that there are interactions between Phenotype i.e., Cfz/Control and biosample, as their patterns intersect each other, in the renal-urinary level. This confirmed that the drug triggered different types of metabolic alterations in each system: corresponding impact in blood and kidneys and reverse impact in urine. Particularly, the levels of metabolites were equivalent increased in plasma and kidney and decreased in urine of Cfz-mice. The leverage and SPE score, Figure 22b, were considered to detect the variables who fit the ASCA-given major patterns: 38 and 2 for the ESI+ and ESI- respectively. In case of MEBA analysis, the flow of metabolites from blood to kidney and their concomitant excretion in urine, was considered as a trajectory, connecting the circulatory and the urinary system. Thus, plasma, kidney, and urine were considered as temporal profiles and the Cfz and Control phenotypes as the different conditions. So, the MEBA time-series was applied to investigate the existence of variables that were significantly influenced by the administration of Cfz in one temporal profile, e.g., metabolite levels increased only in the kidney. Variables were evaluated using the Hotelling-T2 test with values >30 being statistically significant. Only two of

the MEBA proposed variables were also found as differentiating by the other statistical tests i.e., *212.8437_1.23* was a priority variable in both plasma (+) and urine (+), and *631.2701_8.15* in plasma (+).

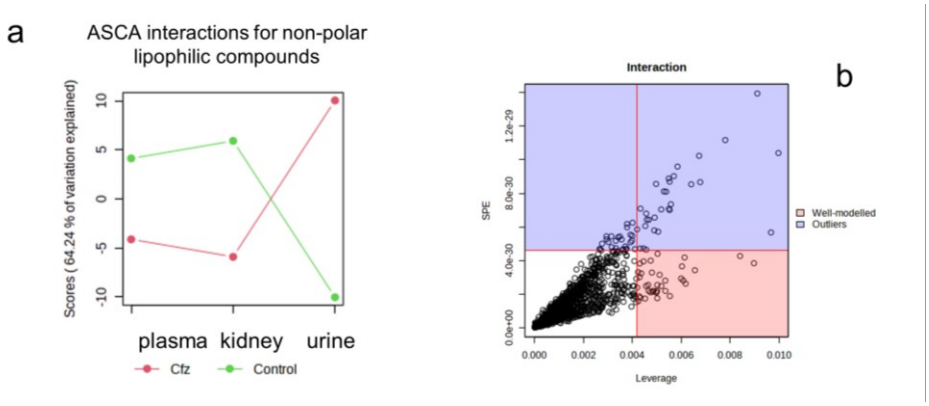

Figure 22 (a) ASCA interactions plot shows that Cfz affects in a similar way the circulation (plasma-kidney) but has opposite effect on urinary system (kidney-urine) (b) Plot showing the variables exhibiting interaction due to their phenotype (Cfz/Control) and the biosample type (plasma, kidney, urine).

5.3.2. Identification

For the identification of the library free protocol, the experimental data, i.e., the pseudoMSMS created by MSDIAL, employing DIA data, were compared with reference MS/MS and in silico MS/MS data, provided by online libraries e.g., HMDB, Figure 23. Both peak-picking methods identified 152 statistically important metabolites, the 57 resulted from the library based approached, Table 8, using the 'suspect-screening list'. The rest resulted from the identification procedure of the unknown features, Table 9.

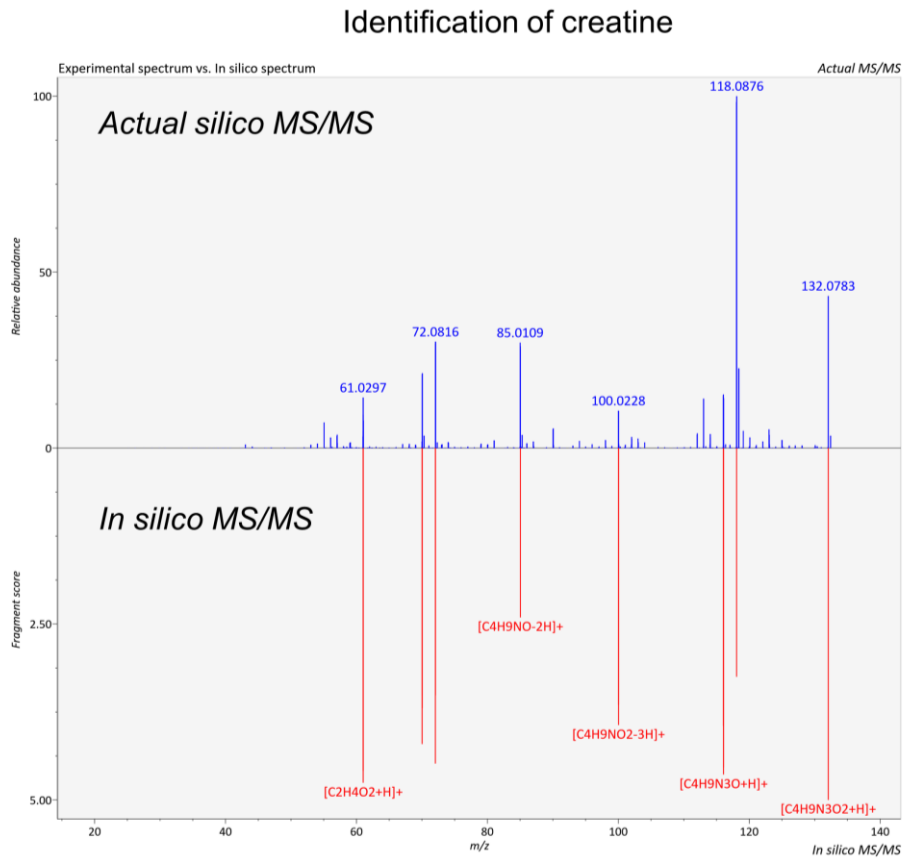

Figure 23 Comparison of the experimental spectrum of the unknown feature with the in silico spectrum of Creatine, using MSFINDER software

Table 8 Metabolites detected using the in-house RPLC-ESI positive metabolites database

Dataset	Metabolite	Cfz	AUC	p.value	VIP	Comment
kidney(-)	D-(+)-Galactose	↓	0.9	0.0306	2	
kidney(-)	D-(+)-Galacturonic acid	↓	0.9	0.0186	2	
kidney(-)	D-Glucuronic acid	↓	0.9	0.0785	2	
kidney(-)	L-Phenylalanine	↑	0.8	0.8056	2	
kidney(-)	Maleic acid	↑	0.8	0.0388	2	
kidney(+)	[(S)-3-Hydroxy-N-methylcoclaurine+NH]	↑	1	0.0044	2	*
kidney(+)	[166.4847_6.96]	↑	0.9	0.0118	2	*
kidney(+)	[166.8851_6.95]	↑	1	0.0037	2	*
kidney(+)	[4-Ethylbenzoic acid +NH]	↑	0.8	0.027	2	*
kidney(+)	[86.3838_6.94]	↑	0.9	0.028	2	*
kidney(+)	[L-Tyrosine -O]	↑	0.8	0.027	2	*
kidney(+)	1-Phenylethylamine	↑	0.9	0.0065	2	
kidney(+)	2-Methylglutaric acid	↓	0.9	0.0191	2	
kidney(+)	3-Methylglutaric acid	↓	0.9	0.0191	2	
kidney(+)	4-Quinolinecarboxylic acid	↓	0.9	0.0144	2	
kidney(+)	Adipic acid	↓	0.9	0.0191	2	
kidney(+)	Cortexolone	↑	0.9	0.0143	2	
kidney(+)	DL-Normetanephrine	↑	0.8	0.027	2	
kidney(+)	Hippuric acid	↓	1	0.0062	2	
kidney(+)	L-Asparagine	↑	0.9	0.0142	2	
kidney(+)	L-Phenylalanine	↑	0.8	0.027	2	
kidney(+)	NG,NG-Dimethylarginine	↑	0.9	0.003	2	
kidney(+)	NG,NG'-Dimethyl-L-arginine	↑	0.9	0.003	2	
kidney(+)	Perillyl alcohol	↓	0.9	0.0196	1	
kidney(+)	Pyridoxamine	↑	0.9	0.0	1.8	
kidney(+)	Quinaldic acid	↓	0.9	0.0	1.6	
plamsa(+)	L-Homoserine	↓	0.8	0.2906	2	
plamsa(+)	L-Methionine	↓	0.8	0.1365	2	
plamsa(+)	L-Threonine	↓	0.8	0.2604	2	
plamsa(+)	N1-Methyl-2-pyridone-5-carboxamide	↑	0.9	0.03	2	
plasma(-)	N-Acetyl-L-leucine	↓	0.9	0.0027	2	
urine(-)	D-(+)-Galactose	↑	0.9	0.0156	1	
urine(-)	Methylmalonic acid	↓	1	0.003	2	
urine(-)	Nonanoate	↓	1	0.0019	2	
urine(+)	(R)-Salsolinol	↓	1.0	0.0	2.0	*
urine(+)	[(R)-Salsolinol+C3H5NOS]	↓	1.0	0.0	2.0	*
urine(+)	[4-Ethylbenzoic acid +NH]	↓	0.9	0.0274	2	*
urine(+)	[L-Tyrosine -O]	↓	0.9	0.0274	2	*
urine(+)	[Pipecolic acid +C4H3N3]	↓	0.9	0.0077	2	*
urine(+)	[p-Synephrine+NH]	↑	0.9	0.0032	2	*
urine(+)	1-Methyladenosine	↑	0.9	0.0221	2	
urine(+)	3-Amino-4-hydroxybenzoic acid	↑	0.8	0.0254	2	

urine(+)	4-COUMARATE	↓	0.8	0.0285	2
urine(+)	4-Ethylbenzoic acid	↓	0.9	0.0108	2
urine(+)	5-Hydroxy-L-tryptophan	↓	0.9	0.0215	2
urine(+)	Argininic acid	↑	1	0.0245	2
urine(+)	DL-Normetanephrine	↓	0.9	0.0274	2
urine(+)	L-Citrulline	↑	1	0.026	2
urine(+)	L-Phenylalanine	↓	0.9	0.0274	2
urine(+)	L-Tyrosine	↓	0.9	0.0074	2
urine(+)	Lumazine	↓	0.9	0.0125	2
urine(+)	L-Valine	↓	0.9	0.0174	2
urine(+)	N(6)-Methyllysine	↑	0.9	0.005	2
urine(+)	N-Acetyl-5-hydroxytryptamine	↓	1	0.0012	2
urine(+)	N-Methyltryptamine	↓	0.9	0.0192	2
urine(+)	Salsolinol	↓	1.0	0.0	2.0
urine(+)	Tyramine	↓	0.9	0.0	1.7

(*) reference in the “Comment-Column” describes those metabolites that were added in the library-based list of peak-picking as Cfz-nephrotoxicity markers, emerged by previous study.

Table 9 List of identified metabolites, resulted from Cfz RPLC untargeted analysis.

Biosample	m/z	RT	Formula	Error (mDa)	Score (MS1)	Regulation (Cfz)	Compound	Score (MS2)	Adduct Type	Library	Reaction
Kidney	205.0991	3.61	C11H11NO2	-0.0025	1	↑	3-Indolepropionic acid	0.93	[M+H]+	One-Reaction/ MCID	[+NH]
Kidney	175.1204	1.29	C7H16NO2	-0.00065	1	↑	Acetylcholine	0.77	[M+H]+	One-Reaction/ MCID	[+CO]
Urine	156.9912	1.72	C4H8O4	-1.4	4.1	↑	(S)-3,4-Dihydroxybutyric acid	7.2	[M-H]-	HMDB (MSFINDER)	-
Urine	308.1059	4.99	C12H21NO4S2	-7	3.7	↑	(S)-Succinyldihydrolipoamide	6.1	[M+H]+	HMDB (MSFINDER)	-
Urine	267.236	13.65	C17H32O2	-3	5.1	↑	10Z-Heptadecenoic acid	7	[M-H]-	HMDB (MSFINDER)	-
Kidney	144.0475	5.51	C9H7NO	-2.3	4	↓	1H-Indole-3-carboxaldehyde	7.4	[M-H]-	HMDB (MSFINDER)	-
Kidney	601.3402	3.31	C27H53O12P	-5.1	3.8	↑	1-Stearoylglycerophosphoinositol	6.3	[M+H]+	HMDB (MSFINDER)	-
Kidney	142.0492	1.31	C6H7NO3	0.7	3.6	↑	2-Aminomuconic acid semialdehyde	6	[M+H]+	HMDB (MSFINDER)	-
Kidney	129.0202	1.2	C5H6O4	-1.5	3.7	↓	2-Hydroxyglutaric acid lactone	6.3	[M-H]-	HMDB (MSFINDER)	-
Kidney	170.0618	3.64	C8H11NO	-3.97609	1	↑	2-Hydroxyphenethylamine	0.77	[M+H]+	One-Reaction/ MCID	[+CO]
Urine	368.2815	10.81	C21H37NO4	-2.2	4	↑	3, 5-Tetradecadiencarnitine	7.05	[M+H]+	HMDB (MSFINDER)	-
Urine	227.1292	4.64	C12H18O4	-1.5	3.8	↓	3,4-Methylenesebacic acid	6.4	[M+H]+	HMDB (MSFINDER)	-
Kidney	245.183	1.26	C12H24O3	-0.00881	1	↑	3-Hydroxydodecanoic acid	0.65	[M+H]+	One-Reaction/ MCID	[+CO]
Urine	219.1172	3.94	C10H18O5	5	3.7	↓	3-Hydroxysebacic acid	6.2	[M+H]+	HMDB (MSFINDER)	-

Urine	161.1088	4.89	C8H16O3	-1.4	3.7	↓	3-Hydroxyvalproic acid	6.69	[M+H]+	HMDB (MSFINDER)	-
Urine	130.0665	5.58	C9H7N	-1.3	4.3	↓	3-Methylene-indolenine	6.5	[M+H]+	HMDB (MSFINDER)	-
Urine	259.0945	2.81	C10H14N2O6	-4	4.1	↓	3-Methyluridine	6.6	[M+H]+	HMDB (MSFINDER)	-
Plasma	129.0926	4.46	C7H12O2	-1.5	4	↓	4-Heptenoic acid	6.9	[M+H]+	HMDB (MSFINDER)	-
Kidney	409.1928	3.61	C11H21N3O5	0.000894	1	↑	5-Acetamidovalerate	0.7	[M+H]+	One-Reaction/ MCID	[+C5H3N5]
Urine	296.067	2.09	C8H14N3O7P	-3	4	↓	5-Aminoimidazole ribonucleotide	6.9	[M+H]+	HMDB (MSFINDER)	-
Kidney	296.0681	1.24	C8H14N3O7P	-3.7	3.4	↑	5-Aminoimidazole ribonucleotide	5.4	[M+H]+	HMDB (MSFINDER)	-
Urine	181.0987	2.51	C9H12N2O2	-1.5	4	↓	5-Hydroxykynurenamine	6.5	[M+H]+	HMDB (MSFINDER)	-
Kidney	181.0879	2.26	C9H12N2O2	10	4.2	↑	5-Hydroxykynurenamine	7.2	[M+H]+	HMDB (MSFINDER)	-
Plasma	181.0874	2.27	C9H12N2O2	9.75	3.2	↑	5-Hydroxykynurenamine	5.2	[M+H]+	HMDB (MSFINDER)	-
Urine	321.045	5.61	C10H15N2O8P	4.3	4	↑	5-Thymidylic acid	6.7	[M-H]-	HMDB (MSFINDER)	-
Kidney	212.0781	1.52	C7H9N5O3	-1.4	3.7	↓	6-Carboxy-5,6,7,8-tetrahydropterin	6.2	[M+H]+	HMDB (MSFINDER)	-
Kidney	315.1123	5.19	C14H14N6O3	7.4	3.7	↓	7,8-Dihydropteroic acid	6.3	[M+H]+	HMDB (MSFINDER)	-
Plasma	157.0848	3.62	C8H12O3	1.1	4.3	↑	8-Hydroxy-5,6-octadienoic acid	67	[M+H]+	HMDB (MSFINDER)	-
Urine	220.065	5.12	C11H9NO4	-4	3.5	↓	8-Methoxykynurenate	5.5	[M+H]+	HMDB (MSFINDER)	-
Kidney	117.0587	3.62	C5H8O3	-4	3.7	↑	Alpha-ketoisovaleric acid	7	[M+H]+	HMDB (MSFINDER)	-
Urine	221.0701	5.49	C7H12N2O6	6	3.5	↓	Aspartyl-Serine	6.02	[M+H]+	HMDB (MSFINDER)	-

Urine	130.062	1.55	C4H9N3O2	0.4	4.7	↑	Beta-Guanidinopropionic acid	7.7	[M-H]-	HMDB (MSFINDER)	-
Kidney	149.0612	3.06	C9H8O2	-1.6	3.7	↑	Cinnamic acid	6.5	[M+H]+	HMDB (MSFINDER)	-
Kidney	132.083	3.62	C4H9N3O2	-5.3	3.9	↑	Creatine	8.2	[M+H]+	HMDB (MSFINDER)	-
Urine	112.0516	2.4	C4H5N3O	-0.8	3.5	↓	Cytosine	6.2	[M+H]+	HMDB (MSFINDER)	-
Plasma	227.0721	9.32	C9H12N2O5	-4.6	0.1	↑	Deoxyuridine	7.9	[M-H]-	HMDB (MSFINDER)	-
Urine	199.0076	4.81	C4H9O7P	-6.3	3.9	↑	D-Erythrose 4-phosphate	6.7	[M-H]-	HMDB (MSFINDER)	-
Kidney	240.1038	2.63	C9H13N5O3	5	4.1	↑	Dihydrobiopterin	7.6	[M+H]+	HMDB (MSFINDER)	-
Urine	433.1156	4.13	C21H20O10	-2.9	3.9	↑	Dihydrodaidzein 7-O-glucuronide	6.9	[M+H]+	HMDB (MSFINDER)	-
Kidney	160.1348	1.61	C8H17NO2	-1.6	4	↑	DL-2-Aminooctanoic acid	6.9	[M+H]+	HMDB (MSFINDER)	-
Plasma	160.1339	1.36	C8H17NO2	0.79	4	↑	DL-2-Aminooctanoic acid	7.23	[M+H]+	HMDB (MSFINDER)	-
Urine	229.0189	4.78	C5H11O8P	-6.2	4.6	↑	D-Ribose 5-phosphate	8.3	[M-H]-	HMDB (MSFINDER)	-
Urine	419.1381	4.96	C21H22O9	-4.4	3.6	↑	Equol 7-O-glucuronide	6.41	[M+H]+	HMDB (MSFINDER)	-
Kidney	276.1033	1.77	C12H13N5O3	5.5	3.5	↓	Ethenodeoxyadenosine	6.2	[M+H]+	HMDB (MSFINDER)	-
Kidney	457.1166	4.26	C17H21N4O9P	-3.9	4.1	↑	Flavin Mononucleotide	7.1	[M+H]+	HMDB (MSFINDER)	-
Kidney	261.148	3.72	C11H20N2O5	-3.4	4	↑	gamma-Glutamylisoleucine	6.8	[M+H]+	HMDB (MSFINDER)	-
Urine	223.1097	3.24	C11H14N2O3	-2	3.9	↓	Glycyl-Phenylalanine	6.26	[M+H]+	HMDB (MSFINDER)	-
Kidney	146.062	3.62	C3H7N3O2	-0.00653	1	↑	Guanidoacetic acid	0.23	[M+H]+	One-Reaction/ MCID	[+CO]

Kidney	118.0666	3.62	C3H7N3O2	-5	3.8	↑	Guanidoacetic acid	6.5	[M+H]+	HMDB (MSFINDER)	-
Urine	140.072	4.36	C6H9N3O	-1.3	3.5	↓	Histidinal	5.8	[M+H]+	HMDB (MSFINDER)	-
Urine	213.0928	7.97	C8H12N4O3	5	4.1	↓	Histidinyl-Glycine	6.6	[M+H]+	HMDB (MSFINDER)	-
Urine	183.0672	1.81	C9H10O4	-1.8	3.9	↓	Homovanillic acid	7.3	[M+H]+	HMDB (MSFINDER)	-
Kidney	351.1743	4.96	C10H19NO5	3.002896	1	↓	Hydroxypropionylcarnitine	0.97	[M+H]+	One-Reaction/ MCID	[+ (C5H5N5 - H2O)]
Kidney	137.4083	2.41	C5H4N4O	-1	4.2	↑	Hypoxanthine	8.4	[M+H]+	HMDB (MSFINDER)	-
Urine	157.0668	6.78	C6H8N2O3	-6	3.4	↓	Imidazolelactic acid	6.2	[M+H]+	HMDB (MSFINDER)	-
Kidney	116.0519	3.54	C8H7N	-2.9	4.1	↑	Indole	8.07	[M-H]-	HMDB (MSFINDER)	-
Kidney	331.1674	3.06	C11H9NO2	-0.00272	0.93	↑	Indoleacrylic acid	0.7	[M+H]+	One-Reaction/ MCID	[+C7H13NO2]
Urine	206.0818	5.16	C11H11NO3	-0.5	4	↓	Indolelactic acid	6.5	[M+H]+	HMDB (MSFINDER)	-
Kidney	222.072	4.44	C11H11NO3	0.003536	1	↓	Indolelactic acid	0.8	[M+H]+	One-Reaction/ MCID	[+O]
Urine	245.0966	6.21	C13H12N2O3	-4.8	3.8	↓	Indolylacryloylglycine	5.8	[M+H]+	HMDB (MSFINDER)	-
Urine	310.0945	3.19	C14H15NO7	-2.5	4.2	↑	Inodxyl glucuronide	7	[M+H]+	HMDB (MSFINDER)	-
Plasma	229.0608	2.76	C11H20N2O3	-0.9	3.6	↓	Isoleucylproline	6.8	[M+Na]+	HMDB (MSFINDER)	-
Kidney	173.1061	1.69	C6H14N4O2	-1.6	5.2	↑	L-Arginine	8.8	[M-H]-	HMDB (MSFINDER)	-
Kidney	869.4803	5.67	C40H72N2O18	-0.5	3.5	↑	Lc3Cer	6	[M+H]+	HMDB (MSFINDER)	-
Kidney	162.114	1.59	C7H15NO3	-1.4	4.1	↓	L-Carnitine	7.4	[M+H]+	HMDB (MSFINDER)	-

Urine	198.0775	4.36	C9H11NO4	-1.4	4.1	↓	L-Dopa	4.34	[M+H]+	HMDB (MSFINDER)	-
Kidney	260.9009	1.18	C11H21N3O4	8	3.8	↓	Leucyl-Glutamine	6.3	[M+H]+	HMDB (MSFINDER)	-
Urine	229.1557	1.96	C11H20N2O3	-0.63	4	↑	Leucylproline	7.2	[M+H]+	HMDB (MSFINDER)	-
Urine	322.1084	7.24	C11H19N3O6S	-0.46	4	↓	L-L-Homoglutathione	6.34	[M+H]+	HMDB (MSFINDER)	-
Kidney	203.0821	3.54	C11H12N2O2	-0.7	4.3	↑	L-Tryptophan	8.2	[M-H]-	HMDB (MSFINDER)	-
Urine	247.1092	4.95	C10H18N2O3S	1.3	3.7	↓	Methionyl-Proline	6.1	[M+H]+	HMDB (MSFINDER)	-
Kidney	264.1019	3.15	C11H15N5O4	0.007764	1	↓	N6-Methyladenosine	0.73	[M+H]+	One-Reaction/ MCID	[-H2O]
Kidney	365.154	4.52	C14H24N2O9	2.5	4.4	↓	N-Acetylmuramoyl-Ala	7.4	[M+H]+	HMDB (MSFINDER)	-
Kidney	265.0966	3.15	C5H9NO4	0.007765	1	↓	N-Acetylserine	0.67	[M+H]+	One-Reaction/ MCID	[+ (C5H5N5 - H2O)]
Urine	175.1494	7.06	C8H18N2O2	-5	3.9	↓	Ne,Ne dimethyllysine	6.41	[M+H]+	HMDB (MSFINDER)	-
Kidney	279.0938	4.01	C11H12N2O4	0.003199	0.9	↓	N'-Formylkynurenine	0.6	[M+H]+	One-Reaction/ MCID	[+C2H2O]
Kidney	131.0508	3.06	C4H6N2O2	-0.00623	1	↑	N-Methylhydantoin	0.6	[M+H]+	One-Reaction/ MCID	[+O]
Kidney	613.1562	3.08	C20H32N6O12S2	3	3.8	↑	Oxidized glutathione	6.9	[M+H]+	HMDB (MSFINDER)	-
Urine	255.5898	13.86	C16H32O2	0.15	4.3	↑	Palmitic acid	8.5	[M-H]-	HMDB (MSFINDER)	-
Urine	181.0662	6.79	C7H8N4O2	5	3.8	↓	Paraxanthine	6.72	[M+H]+	HMDB (MSFINDER)	-
Urine	121.0658	7.97	C8H8O	-0.8	4	↓	Phenylacetaldehyde	6.7	[M+H]+	HMDB (MSFINDER)	-
Urine	165.0564	1.85	C9H8O3	-1.4	3.7	↓	Phenylpyruvic acid	6.5	[M+H]+	HMDB (MSFINDER)	-

Kidney	165.0561	1.88	C9H8O3	-1.9	3.7	↑	Phenylpyruvic acid	6.4	[M+H]+	HMDB (MSFINDER)	-
Urine	304.1777	1.74	C14H25NO6	-2.2	3.9	↓	Pimelylcarnitine	6.74	[M+H]+	HMDB (MSFINDER)	-
Urine	230.1195	7.66	C9H15N3O4	-1.5	3.7	↓	Prolyl-Asparagine	5.85	[M+H]+	HMDB (MSFINDER)	-
Urine	131.1043	5.15	C7H14O2	2.8	3.4	↓	Propyl butyrate	5.8	[M+H]+	HMDB (MSFINDER)	-
Urine	169.0985	2.99	C8H12N2O2	-1.2	3.8	↓	Pyridoxamine	6.9	[M+H]+	HMDB (MSFINDER)	-
Kidney	201.1159	3.53	C10H18O4	-2.7	5.1	↓	R-2-Hydroxy-3-methylbutanoic acid 3-Methylbutanoyl	8.5	[M-H]-	HMDB (MSFINDER)	-
Urine	455.0823	3.44	C13H19N4O12P	-0.8	3.8	↓	SAICAR	6.7	[M+H]+	HMDB (MSFINDER)	-
Plasma	380.258	13.19	C18H38NO5P	-3	4	↑	Sphingosine 1-phosphate	6.8	[M+H]+	HMDB (MSFINDER)	-
Plasma	190.0878	6.55	C8H15NO2S	1.4	4	↓	S-Prenyl-L-cysteine	6.2	[M+H]+	HMDB (MSFINDER)	-
Kidney	99.0089	1.56	C4H4O3	-0.14	3.9	↓	Succinic anhydride	6.29	[M-H]-	HMDB (MSFINDER)	-
Urine	391.176	7.67	C22H22N4O3	0.4	3.9	↓	Tryptophyl-Tryptophan	6	[M+H]+	HMDB (MSFINDER)	-
Urine	113.0353	2.61	C4H4N2O2	-0.8	3.6	↓	Uracil	6.3	[M+H]+	HMDB (MSFINDER)	-
Kidney	113.0361	1.93	C4H4N2O2	-1	4.2	↑	Uracil	7.12	[M+H]+	HMDB (MSFINDER)	-
Kidney	147.1143	1.3	C5H10N2O4	-1.5	3.8	↑	Ureidoisobutyric acid	7	[M+H]+	HMDB (MSFINDER)	-
Urine	139.0407	7.18	C6H6N2O2	-2	3.4	↓	Urocanic acid	6.5	[M+H]+	HMDB (MSFINDER)	-

5.4. Discussion

This study aimed to enrich the existing knowledge on Cfz induced nephrotoxicity, by the metabolomics prospective, performing RP chromatographic analysis, as a complementary step to Barla et. al., published workflow[1]. Metabolomics deals with thousands of compounds, therefore, the combination of different separation methodologies offers higher opportunities for 'good performance' for a larger number of metabolites. In other words, compounds that co-eluted in RP, can be separated using HILIC. The current RP-based study confirmed the HILIC-based observations regarding the strong impact of Cfz in the circulatory and urinary systems and furthermore provided new evidence. The first important notice is that the non-polar metabolome (RP data) provided more clear classification trends, concerning PCA, between Cfz and Control mice in all datasets. This indicated that Cfz affects the non-polar metabolome in a more intense way. Also, the RP dataset led to the identification of more than 90 significantly differentiating metabolites, while HILIC protocol identified 53. Thus, RP is considered as a more informative approach.

In the current project, library-free and library-based methodologies were combined for metabolites detection. It is worth mentioning that the metabolites provided by the library-based search, red-marked loadings at Figure 24, showed limited discriminating ability compared to the corresponding of the library-free protocol. The library-based approach showed limitations regarding the interpretation of Cfz impact, as more than 700 statistically significant features remained unknown. Specifically, 17% of the overall important features were finally identified, with 6% being identified by the library-based protocol. The library-free methods are more informative, as they are not limited on already known compounds, i.e., usually amino acids, carnitines, and other naturally occurring compounds. In addition, they focus on the differentiating statistically important metabolic evidence, either known or not, giving the chance to record features of higher specificity for the examined condition. In the current case thirteen of the identified metabolites were detected as products of metabolites' metabolism, e.g., as acyl-derivatives, generated by enzymatic actions on metabolites.

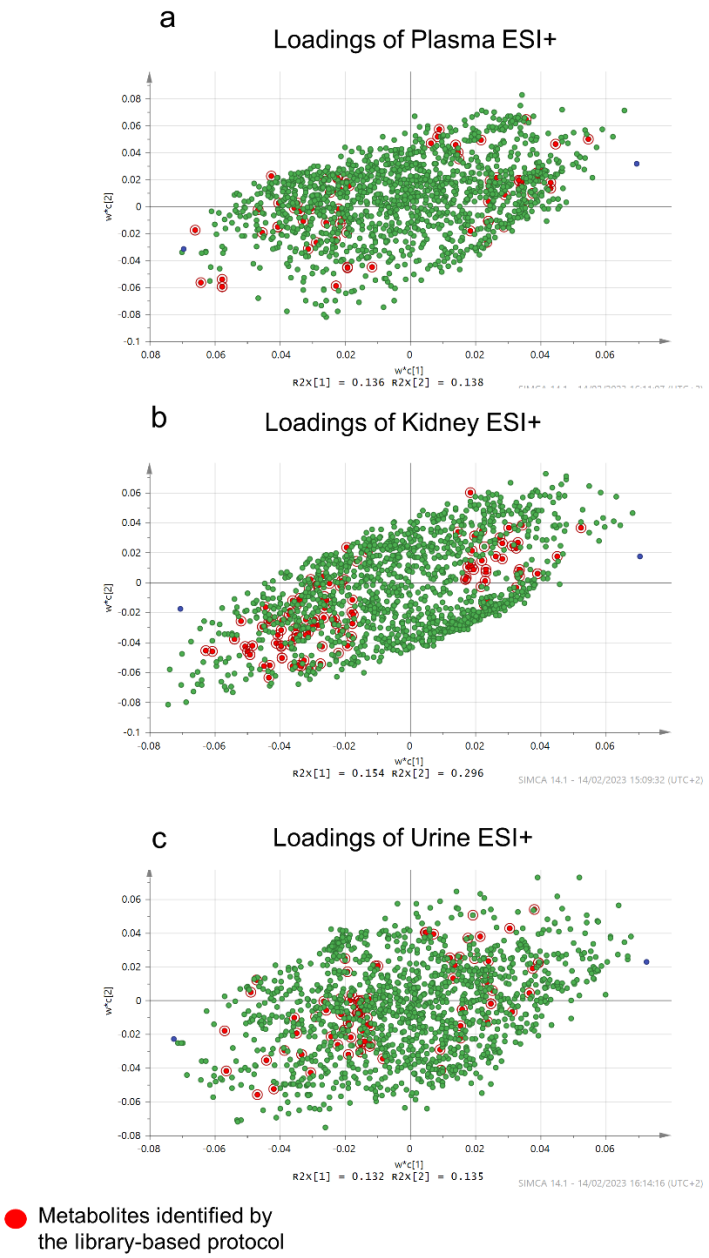

Figure 24 Plot of loadings of plasma ESI+ (a); kidney ESI+ (b); and urine ESI+(c). The red marked points represent the metabolites detected from the library-based approach.

5.4.1. Patterns of Cfz induced metabolic regulation.

It was observed that for the Cfz-treated mice, 65% of the metabolites increased their levels in plasma and kidneys and the 65% decreased their levels in urine excretes. The later suggested two opposite phenomena: i) increase of metabolites in the blood circulation and ii) decrease of them in excretes. However, this observation is in accordance with what we have previously

shown. Cfz leads to water and salt retention via mineralocorticoid receptor (MR) *in vivo* activation to the kidneys, and in patients with Cfz-related renal adverse events as well. This patho-mechanism was further supported, as Eplerenone administration, a clinically applicable MR blocker, reversed Cfz's nephrotoxicity *in vivo* by Efentakis et. al.[51]. The current data support these findings and the increase in renal and circulatory metabolites can be attributed to impairment of renal reuptake mechanisms. Aiming to find patterns that are the representative ones, for the majority of the metabolites, and in contrast to the previous applied univariate approach, the current study developed two multivariate approaches, i.e., ASCA and MEBA. So, ASCA enable the detection of interactions between the drug and the circulatory/urinary system as well, besides estimating the significance of the main effects. Additionally, the MEBA step of the proposed workflow, aimed to distinguish variables that exhibit significant dysregulation only in a single point, by assigning this "behaviour" to their temporal profile. However, MEBA did not afford significant number of *single point dysregulated metabolites*, in contrary to the univariate approach that was applied in the HILIC data, Figure 25. A significant number of statistical important polar metabolites (from the HILIC-data) were found accumulated only in the kidneys, while the majority of non-polar lipophilic substances (from RP-data) did not show a single dysregulation peak in the temporal profiles, confirming the ASCA findings.

Kidneys perform filtration, reabsorption, secretion and, metabolism; thus renal dysfunction affects the circulating metabolites in different, even controversial ways[54]. The abnormal accumulation of specific metabolites could either result by the kidneys inability to catabolize substances, by extensive renal biosynthesis, or by abnormalities in reabsorption/excretion. Increased metabolite levels in blood could be caused by their limited renal clearance as well. Though, it is interesting that the polar metabolites (HILIC data), were mainly accumulated in kidneys, while their lipophilic counterparts were increased in both blood and the kidneys. This betrays perturbation to the metabolic transportation pathways. The existing data are in accordance with these two facts: i) Cfz affects renal osmolarity leading to alterations of the of ions' exchange routes in kidneys, ii) Cfz activates the MR, leading to the expression of proteins that regulate the ion/water transportation. Potentially, the

later inhibits the renal transformation of lipophilic molecules into polar derivatives, that are easily excreted by the glomerular filtration barrier. This obstructs the normal renal clearance, explaining the corresponding elevated levels of non-polar metabolites in blood.

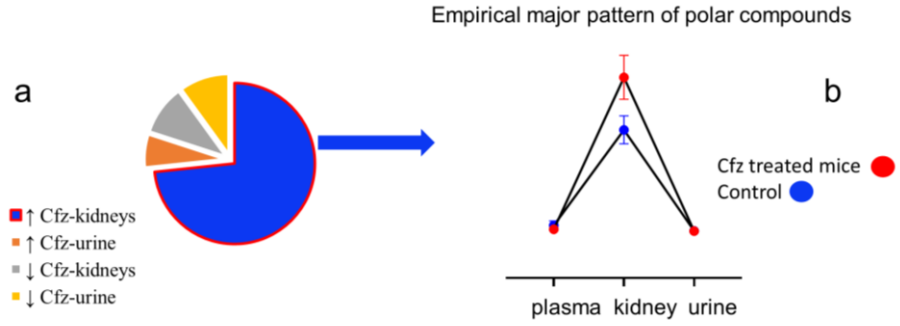

Figure 25 (a) Description of the regulation of polar metabolites (HILIC data) in the Cfz-treated mice; (b) empirical major pattern of polar metabolites (HILIC data)

5.4.2. Biomarkers of Cfz nephrotoxicity

The identified metabolites were searched in the database of the European Uremic Toxin Group (https://database.uremic-toxins.org/soluteList.php) and creatine (↑ Cfz-kidneys), hypoxanthine (↑ Cfz-kidneys), N6-methyladenosine (↓ Cfz-kidney), argininic acid (↑ Cfz-urine) and, 1-methyladenosine (↑ Cfz-urine), were identified as uremic biomarkers, and except of n6-methyladenosine, the levels of the rest were found increased as in uremic conditions. Regarding the rest of the metabolites:

Eleven metabolites participate in tryptophan (Trp) degradation pathway, i.e., tryptophan, 3-indole-propionic acid, 1H-indole-3-carboxaldehyde, 2-aminomuconic acid semialdehyde, 5-hydroxykynurenamine, indole, indoleacrylic acid, indolelactic acid, N'-formylkynurenine.

Four metabolites are hydroxy fatty acids (HFAs) i.e., (S)-3,4-dihydroxybutyric acid, 3-hydroxydodecanoic acid, 4-8-hydroxy-5,6-octadienoic acid, were up-regulated whereas 3-hydroxyvalproic acid, was downregulated.

Three metabolites are FA-derivatives, i.e., 3,5-tetradecadiencarnitine, hydroxypropionylcarnitine and 3-hydroxyglutaric acid lactone.

Three metabolites are medium and long chain fatty acids, i.e., 10Z-heptadecenoic acid and palmitic acid found with increased levels and two, i.e., heptenoic acid, 7,8-dihydropteroic acid and 3,4-methylenesebacic acid were found decreased.

5.4.3. Dysregulation of fatty acids oxidation.

Cfz treated mice showed increased levels of HFAs and derivatives, showing the Cfz induced dysregulation of FA metabolism. It is known that the defective FA metabolism, i.e., uptake, formation and degradation provoke severe dysregulation in renal function (25,26). Kidneys' filtration-reabsorption operation requires high demand of energy, provided mainly by fatty acids oxidation (FAO) [55],[56]. FAO is accomplished by the mitochondrial β-oxidation, by the peroxisome β-oxidation or by the microsomal ω- oxidation. Short and medium FA are preferably metabolized by mitochondria, whereas the long chain FA by peroxisomes. Peroxisomes break-down the FA chain, and their shorter products are moved into mitochondria to complete the FAO procedure [55],[57]. The 3-HFAs are substrates for the peroxisomal β-oxidation, and their increased levels may enhance this procedure, leading to augmented formation of reactive oxygen species byproducts. The inhibition of PPARa expression also limits the production of short acyl-CoA, that are metabolized in the mitochondria to generate ATP. The latter case indicates mitochondrial dysfunction. These two scenarios could be described by the observed increase of HFAs in kidneys. The peroxisomal β-oxidation is regulated by PPARa receptor, which is expressed in kidneys and its dysregulations has been related to acute kidney injury (AKI) [55],[57]. Previous results of molecular analysis had proven that Cfz prohibits the phosphorylation of adenosine monophosphate-activated protein kinase (AMPK) [53], making this connection the most prevailing scenario. AMPK regulates the levels of the circulating free FAs, by i) activating the FAO procedure, as AMPK elevates the activity of carnitine palmitoyltransferase-1 (CPT-1); ii) inhibiting lipolysis and lipogenesis [58]. Hence, in the current case, the increase of HFAs levels betray limitations in CPT-1 activity.

5.4.4. Dysregulated metabolic pathways.

The identified metabolites were subjected to pathway analysis, revealing significant alterations in the pathways of aromatic amino acids' (AAAs)

regulation. Specifically, the affected pathways were: i) the AAAs' biosynthesis; ii) the phenylalanine (Phe) metabolism; and iii) the tryptophan (Trp) metabolism, Figure 26. It is important to note that the pathways of AAAs' biosynthesis, and Phe metabolism were dysregulated in both kidneys and urine of Cfz treated mice, but in opposite directions, i.e., upregulated in kidneys and downregulated in urine, Figure 7. Trp metabolism was only altered in the renal level, showing increase for the Cfz treated mice.

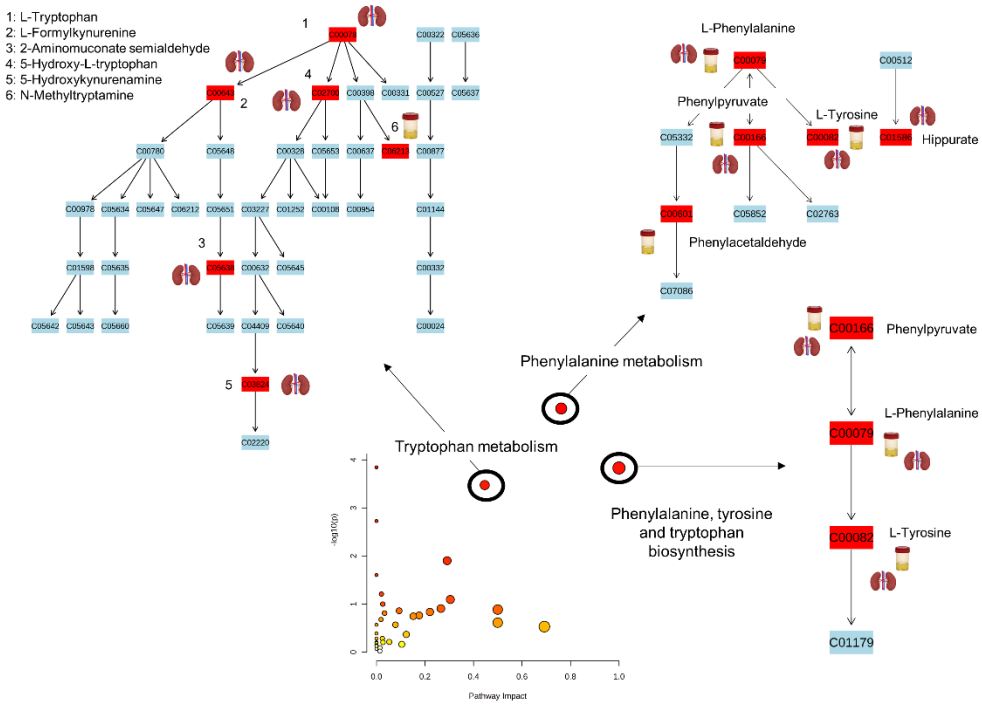

Figure 26 Summary of the altered metabolic pathways resulted by the pathway analysis using MetaboAnalys 5.0.

5.4.4.1. Dysregulation of biosynthesis and metabolism of phenylalanine

The regulation of AAAs (Phe, Tyr, Trp) is crucial for the normal kidney function [59] . The levels of the circulatory AAAs, except of Trp, seem to be negatively correlated to the glomerular filtration rate (GFR) as their levels are increased in limited GFR patients, showing limited kidney uptake [59]. However, in the current case the opposite condition is observed. The levels of Phe, Tyr and Trp unchanged in the blood of the Cfz treated mice, whereas they are found increased in the kidneys and decreased in the urine. Despite their elevated

renal levels, their catabolic products (phenylethylamine, 2-hydroxyphenylactate, etc.) were not altered in the kidneys of Cfz treated mice, indicating inhibition of AAAs degradation. On the other hand, AAAs and their catabolic products were both decreased in urine, confirming that the AAAs are retained by the kidney, Figure 27.

In addition, AAAs are precursors of uremic toxins e.g., indoxyl-sulphate[60]. Thus, their elevation combined with the increased uremic toxins, e.g., hypoxanthine (↑ Cfz-kidneys), argininic acid (↑ Cfz-urine) and, 1-methyladenosine (↑ Cfz-urine), strengthens the suggestion the establishment of early uremic conditions due to the administration of the drug.

The decrease of L-Dopa and tyramine in urine is an interesting finding, as they are precursors of dopamine. The decarboxylation of circulatory L-dopa is the main source of dopamine in the kidneys, and the renal dopamine represents the main source of the urine dopamine. It was hypothesised that the decrease of urine excreted L-dopa, reflects the corresponding decrease of renal L-Dopa and consequently impairment of the renal dopamine biosynthesis. As the compound is a regulator of water and electrolyte balance, facilitating their excretion[61], the renal dopamine was further investigated. Despite that the compound had not passed the statistical thresholds (AUC>0.9, VIP>1.5, FDR p.value<0.05), and therefore was not considered as a potential biomarker, its levels were found decreased in kidney of Cfz treated mice (p.value=0.046), but not altered in urine (p.value=0.34). This result could indicate the biochemical background of the increased systolic blood pressure that occurs as an adverse effect of Cfz [51], due to decreased vasodilation in the kidneys as well as impairment of the hormonal regulation.

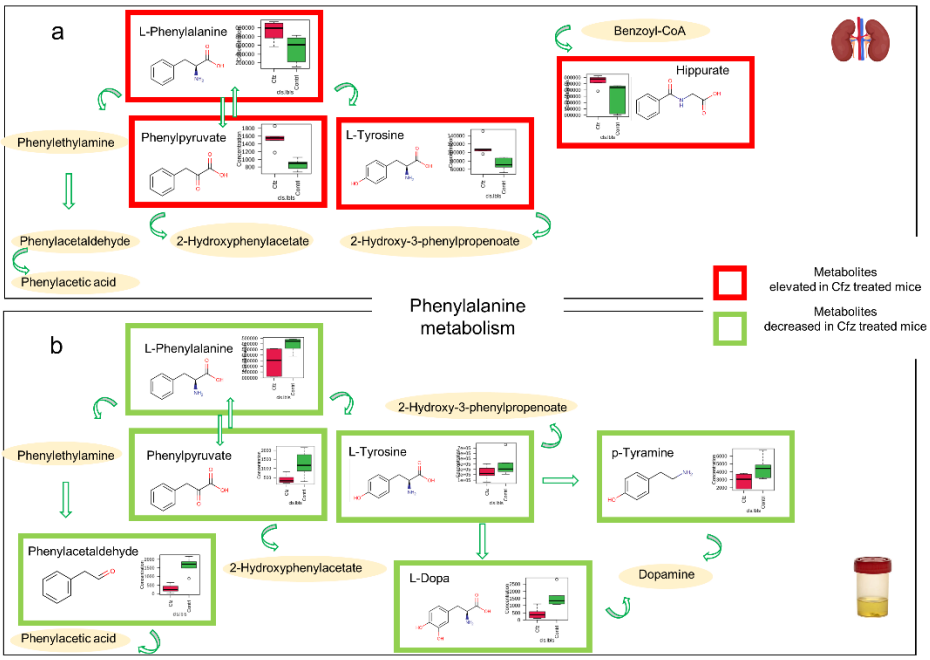

Figure 27 Summary of Phe metabolism in kidney (a) and urine (b). The metabolites included in red border were upregulated and those included in green border were down regulated in Cfz-treated mice. The metabolites described only by their name were not affected by their name were not affected by the drug.

5.4.4.2 Dysregulation of tryptophan metabolism.

There are three pathways of Trp catabolism: the serotonin pathway, the kynurenine pathway, and the indole pyruvate pathway, which involves microbial metabolism [62]. These pathways lead to either products that maintain their indole ring, e.g., serotonin, indoxyl sulphate, and indolelactate, or products that break the indole ring, producing kynurenine derivatives [63]. In the current study, the levels of Trp increased in kidney of Cfz treated mice whereas the kynurenine-derivative, i.e., 5-hydroxykynurenamine, was increased in plasma and kidneys, whereas the 5-Hydroxy-L-Tryptophan was found significantly increased in kidneys of Cfz treated mice. Regarding the identified indole-derivatives, no specific regulation pattern can be shown between Cfz and Control groups, as 3-indolepropionic acid, indole and indoleacrylic acid are elevated, whereas the 1H-indole-3-carboxaldehyde, 3-methylene-indolenine, and indole lactic acid are decreased. Indole lactic acid was detected in both kidney and urine of Cfz treated mice appears as downregulated. Trp downregulation suggests increase of its degradation rate via the kynurenine pathway, condition linked to developed stages of renal disease [64]. The final product of this route is NAD, however, neuroactive compounds also appear as

intermediate products, responsible for the formation of free radicals [64]. The indole-products of Trp, are mainly formatted by microbiota and contribute either positively or negatively, as anti-inflammation factors, or worsen the kidney-dysfunction conditions, as happens with the indoxyl-sulfate, the final product of indole pathway[62]. Most of the metabolites involved in Trp metabolism were detected in kidneys. The significant dysregulation of Trp catabolism was statistically confirmed by pathway analysis, showing an impact value of 0.41 (pathway topology analysis) and an FDR p.value lower than 0.002 (enrichment analysis), Figure 28.

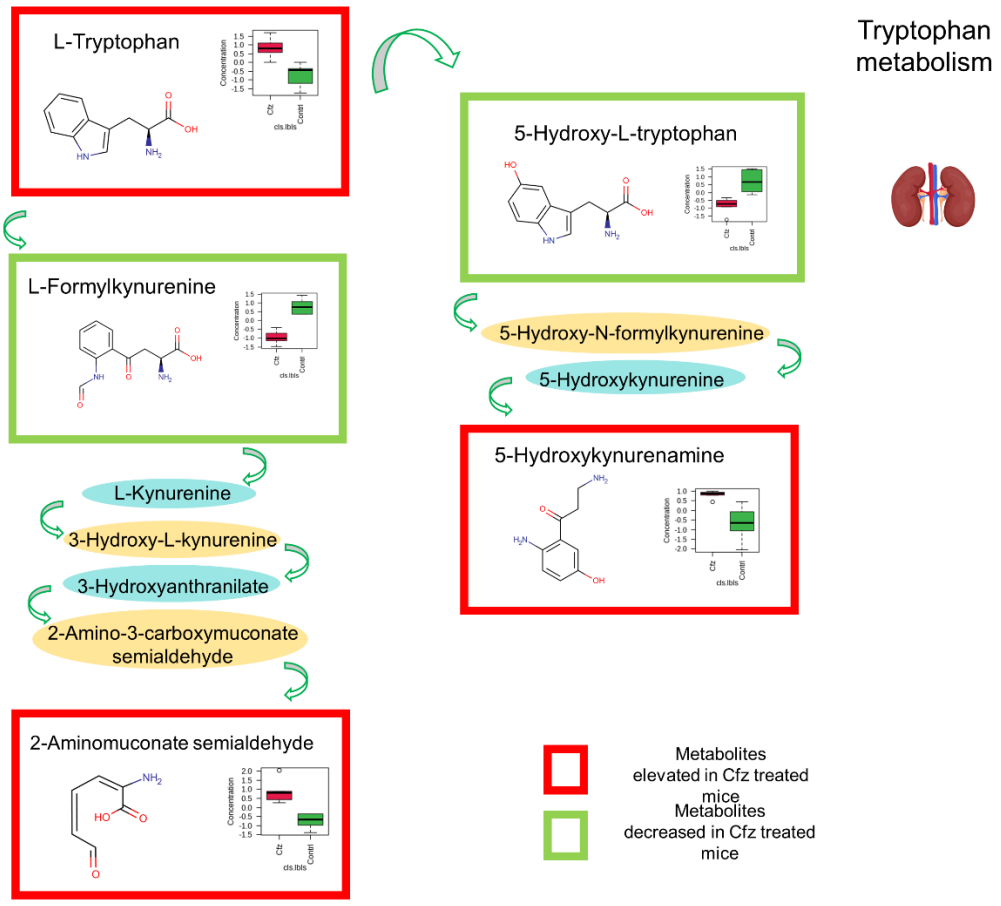

Figure 28 Summary of the detected alterations occurring in Trp degradation pathway in renal level. The metabolites included in red border were upregulated and those included in green border were down regulated in Cfz-treated mice. The metabolites described only by their name were not affected by the drug.

5.4.4.3. Kynurenine sub-pathway of Trp catabolism

The kynurenine pathway has been shown to play an important role in the progression of kidney related disease [65]. The renal dysregulation as a path of the cardiovascular disease development is also closely related to the Trp

metabolism through the kynurenine pathway [66]. A rather neglected metabolite of the pathway that seems to play a significant role to inflammatory response is 5-Hydroxykynurenamine (5-HKA). The molecule has recently emerged as the third Trp metabolism biogenic amine not involving IDO-1[17]. In the current study 5-HKA was increased in plasma and kidney, and slightly decreased in urine of Cfz treated mice, as shown in Figure 13. Its AUC values were 0.83, 0.96, 0.86 for plasma, kidney, and urine, indicating a strong correlation with the development of nephrotoxicity. There is limited literature regarding this compound: 5-HKA is a product of 5-hydroxy-tryptophan (5-HTrp), formed under the influence of the indoleamine 2,3-dioxygenase [67]. 5-HKA has been reported as serotonin receptor agonist in soft tissues, inducing analogous but not so intense contractile response [68]. Thus, the increased levels of 5-HKA, increase the renal perfusion pressure as the serotonin does, prohibiting the normal procedure of renal autoregulation [69]. On the other hand, 5-HKA has been found to ameliorate kidney inflammation through reducing the levels of inflammatory cytokines[70] whereas treatment with this substance has shown improvement of proteinuria and nitrogen balance. Under this notion, the increase of 5-HKA in the current case could indicate the existence of inflammatory condition. Taking under consideration that renal perfusion in the microvasculature of the glomeruli is a key determinant of blood pressure and since we have previously shown that Cfz induces increased systolic blood pressure via MR activation and subsequent water/salt retention both in vivo and in patients, the increase in 5-HKA might be correlated with these phenomena [51].

5.4.4.4 Indole-pyruvate sub-pathway of Trp catabolism

Indole, indoleacrylic acid (IAcr), indole propionic acid (IPA) and indole-3-lactic acid (ILA) are microbial metabolites of Trp, participating in the indole-pyruvate pathway. It is claimed that Trp's indole-derivatives are beneficial for the intestinal and the systematic homeostasis [71]. Kidney dysfunction dysregulates the microbial metabolism leading to further homeostasis issues. Particularly the abovementioned indole-compounds interact with the aryl hydrocarbon receptor (AhR) and pregnane X receptor (PXR), prohibiting the inflammation [62]. Interestingly, in the current study the indole-compounds were

found increased in kidneys of Cfz treated mice, except of ILA that was found downregulated in the kidneys and urine of the same mice. ILA, is produced by gut microbiota and specifically from the species of *Bifidobacteria* [71]. ILA exhibits anti-inflammatory action, by inhibiting the production and the migration of pro-inflammatory cytokines [72]. In previous studies, the compound has been found increased in urine samples of patients suffering from septic acute kidney damage [73] and liver injury [74]. It is worth mentioning that according to Madella et. al., ILA, in contrast to IPA, indole and IAcr, presents the higher degree of interaction with AhR, whereas none of them interact strong with PRX [62]. This suggests that the decreased levels of ILA, therefore its reduced anti-inflammatory activity, cannot be counterbalanced by the increased levels of the other indole-derivatives that could inhibit nephrotoxicity occurrence.

5.5. Materials and Methods

The sample preparation is included in the Supplementary Material, §2.1 [52]. The reversed phase (RP) chromatographic separation was performed using an Acclaim RSL C18 Column, (2.1 x 100 mm, 2.2 μm, Thermo Fischer Scientific, Dreieich, Germany). The temperature has been kept at 30 ^{o}C, throughout all experiments, with the aid of a software-controlled oven. The mobile phases were (A) aq. Methanol, (95-5, v/v) and; (B) Methanol, both containing 5 mM Ammonium Formate. For the positive ionization the mobile phases were acidified with 0.01% Formic Acid (v/v). The gradient was set as ramp escalation of B, from 0% to 100% in 14 min. The flow was set to 0.2 mL/min and the injection volume to 5μL. A Dionex UltiMate 3000 RSLC (Thermo Fischer Scientific, Dreieich, Germany) UPLC chromatographic system of coupled to a Maxis Impact QTOF mass spectrometer (Bruker Daltonics, Bremen, Germany) with an Apollo ESI source, was used for the analysis. The ESI temperature was set at 200 ^{o}C and the capillary voltage was 2.5 kV for both ionization modes. Data independent acquisition (DIA) was employed obtaining both low-collision energy (CE), 5V, and high-CE, 24-36V (ramp), MS data.

Library-free and library-based methodologies were used for metabolites detection. The library-based approach used a list of known metabolites obtained from two sources: i) an in-house database (DB), (ESI+ and ESI-) developed using the MSMLS Mass Spectrometry Metabolite Library of

Standards, IROA Technologies LLC (Sigma Aldrich, Steinheim, Germany); and ii) a list of metabolites already linked to Cfz nephrotoxicity known by previous study [52]. Regarding the DB metabolites, the information of the retention time and the fragmentation was known, whereas regarding the Cfz-nephrotoxicity metabolites, only the m/z of the precursor ion and the adduct ions were known, and thus, the identification was based on this and on their isotopic profile as well. The software TASQ Client 2.1 (Bruker Daltonics, Bremen, Germany). The untargeted peak-picking workflow employed MSDIAL 4.9.2 software, which was used for the generation of pseudo-MSMS[75] files from the DIA data as well.. The peak-lists of the library-free and library-based protocols were concatenated forming a unique table; the duplicates were removed; and QC-based signal correction was performed, using the QCRFSC algorithm of the StatTarget [76] GUI, in the R statistical language environment.

Regarding the statistical analysis, the univariate tests were performed with MetaboAnalyst 5.0 online software, using the modules: i) “Biomarker Analysis” for univariate ROC curve analysis [77] and ii) “Statistical Analysis (one factor)” for FDR corrected t-testing. In both cases the data were log-transformed and auto-scaled. MetaboAnalyst 5.0 was also used for pathway analysis, using the module “Pathway Analysis (targeted)”. Furthermore, SIMCA 14.1 (MKS Umetrics, Uppsala, Sweden) software was used for the development of PCA and PLS-DA classification models. The raw data were scaled by Unit Variance, UV. The resulting PLS-DA models were evaluated by test of permutation testing, miss-classification test, and models’ ROC curve analysis.

Post-hoc analysis employed the MetaboAnalyst 5.0 module “Statistical Analysis (metadata table)” to perform ANOVA-simultaneous component analysis (ASCA) [78] and multivariate empirical Bayes statistics (MEBA) [79]. For doing this, the plasma, kidney, and urine datasets, of each ESI mode, were concatenated in one matrix, log-transformed and auto-scaled. The phenotype (Cfz vs Control) and the biosample type (plasma, kidney, urine), were specified as the metadata factors.

Finally, for the identification of the unknown features, the (mz_rt) were prioritized by their VIP scores (>1.5), their AUC value (>0.7), and their FDR t-

test p value (< 0.05). MS-FINDER 3.52 [80] and the online databases HMDB (Human), Urine (Human), Serum (Human) and LipidMAPS were used to identify the compounds. The m/z range was set at 100-1000 Da; the m/z tolerance of MS1 at 5 mDa; m/z tolerance of MS2 at 10mDa; fragments relative abundance cutoff was 20%, and the isotopic tolerance was 50%. For compounds annotation both the Formula Finder and Structure Finder algorithms were employed. The MyCompoundID [81],[37] (http://www.mycompoundid.org/mycompoundid_IsoMS/) online library was also used to annotate unknown variables of kidney(+) dataset, using the MSMS search of 1 reaction.

5.6. Conclusions

This study aimed to provide further evidence on the biochemical background of Cfz induced nephrotoxicity, emphasizing in the non-polar metabolites. The metabolomics analysis identified 93 metabolites, 5 of which belong to the class of the uremic toxins. The post-hoc analysis revealed interactions between the Cfz and the circulatory/urinary system, i.e., the levels of metabolites were increased in plasma and kidneys but decreased in urine of Cfz group. This finding was controversial to previous results based on polar metabolites, that shown local increase of the polar metabolites in kidneys, and non-significant alteration detected in plasma level. In the current case, the non-polar metabolites were similarly increased in plasma and kidneys, indicating their limited renal clearance. We found increased levels of HFAs in kidneys of Cfz-treated mice, indicating dysregulation of FAO and mitochondrial dysfunction. The novel founding was that Cfz imposed alterations in AAAs biosynthesis and degradation. Phe metabolism was impacted in kidney and urine level but in opposite ways (increased in kidney, decreased in urine), showing limited excretion of those metabolites in urine. Also, the decreased levels of dopamine in kidney of Cfz-treated mice, implies its limited renal biosynthesis, fact that could explain the decreased vasodilation in the kidneys, leading to increased systolic blood pressure.

6. Chapter 6

Metabolomics exceed the challenge of confounders presenting early biochemical evidence for the cardiotoxicity risk assessment in a cohort of oncology underaged patients.

6.1. Abstract.

Cardiotoxicity (CT) is an adverse effect of chemotherapy that occurs in adults and children. So far, genomics has linked the possibility of CT with specific gene polymorphism. This study used metabolomics as a complementary approach, aiming to find association between the metabolic profile of patients before they undergo chemotherapy and, the risk to display CT during their treatment.: A UPLC-ESI-QTOF protocol was developed for the samples' analysis. The plasma samples were obtained from 89 underaged patients, before they start chemotherapy, and the metabolomics data were combined with the *a posteriori* knowledge of CT expression in a post-hoc analysis. KODAMA, OPLS-DA, BORUTA, and FDR t-test were used as complement classification and variables selection methodologies to point out metabolites with an increased probability of being CT-related. An empirical scale of confidence for DIA identification was also developed. All classification models succeeded in separating the patients of CT-risk. The identified metabolites suggested metabolic differentiations in the pathways of fatty acids and amino acids regulations, that are involved to the general pathway of cardiac energy metabolism. The metabolites hydroxy-9-hexadecenoylcarnitine, glutamylalanine and 2-hydroxy-3-methylpentanoic acid showed AUC scores > 0.85. The final results provided novel evidence demonstrating the existence of early detectable metabolic markers related to the risk of CT occurrence, suggesting that chemotherapy is not the sole reason for CT expression. Probably, chemotherapy triggers preexisting cardiac function abnormalities, accelerating the establishment of cardiac dysfunction incidents.

6.2. Introduction

Cardiotoxicity (CT) is a common adverse effect related to the majority of the antineoplastic drugs. According to the European Society of Cardiology the term *cardiac dysfunction* or *cardiotoxicity*, is attributed to any heart injury related to cancer treatment or cancer itself. The events may be acute, subacute, or

delayed depending on the time of appearance and can be categorized as irreversible or reversible as well [82],[83]. As it is generally suggested, CT onset may depend on the class of the antineoplastic drug, dose, route, duration of treatment, and other individual risk factors [82] and occurs in both adults and children[84]. It is worth noticing that only a percentage of chemotherapy treated patients express a CT event. The CT percentage of adult patients treated with anthracyclines fluctuates from 2 to 48%, depending on the dose[82], whereas 18% of children oncology patients express acute or early episodes[85]. These rates imply that CT is caused not only by the administrated antineoplastic drugs, but other personalized factors also contribute to the condition. So far, these personalized factors have been examined by genomics, in both adults and children cohorts, showing specific gene polymorphisms related to CT risk[86]. As metabolomics is a complementary approach, it could significantly contribute to the investigation of the pre-existing CT trends and support the attempt for the CT risk assessment. Until now, metabolomics have provided important knowledge on the metabolic alterations occurring in CT conditions, improving our comprehension on the disease's biochemical mechanisms [87]. Though, the metabolomics investigation of oncology patients before their submission to any therapeutic intervention, will point out the latent triggering causes of CT expression and could potentially support the early risk assessment.

This study represents a pioneering effort in utilizing metabolomics to examine the association between the metabolic profile of patients before they undergo chemotherapy and the risk to display CT during their treatment. Therefore, a metabolomics protocol was applied in plasma samples of oncology pediatric patients before their submission to antineoplastic treatment, day=0. Afterwards the children followed the proper therapeutic protocol and some of them expressed CT. This *a posteriori* knowledge was combined with the metabolomics data in a post-hoc analysis. The ultimate aim was to anticipate the underlying metabolic background of CT and facilitate the identification of prognostic metabolites thereof.

6.2.1. The analytical challenge

The current study involved 89 children pediatric patients, male and female, in the age range of 0-16 years. These children were diagnosed with several types of hematologic malignancies and solid tumors i.e., acute lymphoblastic leukemia, ALL, acute myeloid leukemia, AML, lymphoma, soft tissue sarcoma, bone tumors, brain tumors, etc. Thus, the study design encompasses the influences of several confounding factors, such as genotype, sex, age, and cancer type, on the children's metabolome. To address the high uncertainty and enhance the reliability of the obtained results, a novel approach was implemented, which involved the combination of four diverse chemometric philosophies for variable extraction:

- Assembly and evaluation of random classification models (knowledge discovery by accuracy maximization, KODAMA) [47];
- Supervised classification based on data reduction (ortho-partial least square discriminant analysis, OPLS-DA);
- Classification using subsets of all variables (BORUTA) [46];
- Classical univariate approach (*t*-test).

A Bayesian logic underlies this notion. The posterior probability is refined by any of the priors if the variables are identified by two or more statistical tests.

6.3. Results

6.3.1. Peak Picking & Signal correction

QC-based and Internal-Standard-based techniques were combined to correct the signal drift, as shown in Figure 29. PCA models were used to assess the efficiency of the signal-correction step by checking QCs reproducibility, and to discover any outliers that had to be excluded.

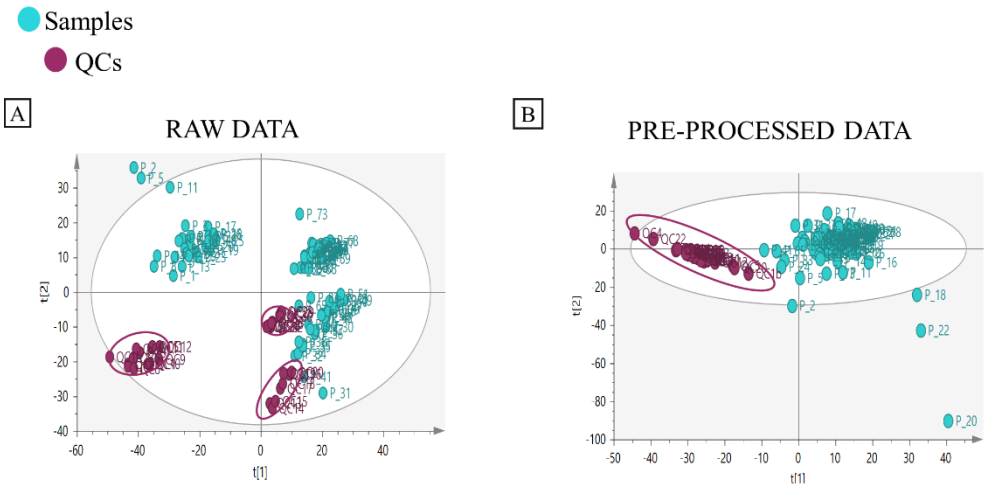

Figure 29 Graphical representation of the signal correction results: (A) PCA scores plot of raw data of QCs and samples (B) PCA scores plot of pre-processed data of QCs and samples.

6.3.2 Statistical analysis results

6.3.2.1. KODAMA

KODAMA is a learning algorithm that creates random classification clusters driven by a Monte Carlo procedure that maximizes the prediction ability. The semisupervised module of KODAMA was employed, forcing the CT-risk patients to belong to the same class during the algorithm's remodelling subroutine. The topological representation of KODAMA semi-supervised classification is shown in Figure 30. The procedure is semi-supervised as the No-risk patients were not forced to belong in the same group, (decreased dissimilarity), neither to be separated from the CT-risk patients (increased dissimilarity between groups).

The results of KODAMA showed, for the first time, early metabolic evidence of CT risk, as the algorithm managed to group the patients based on the risk to express CT. KODAMA employs the Kruskal–Wallis t-test to highlight the discriminating variables (those with log(P value)> 3 were considered important).

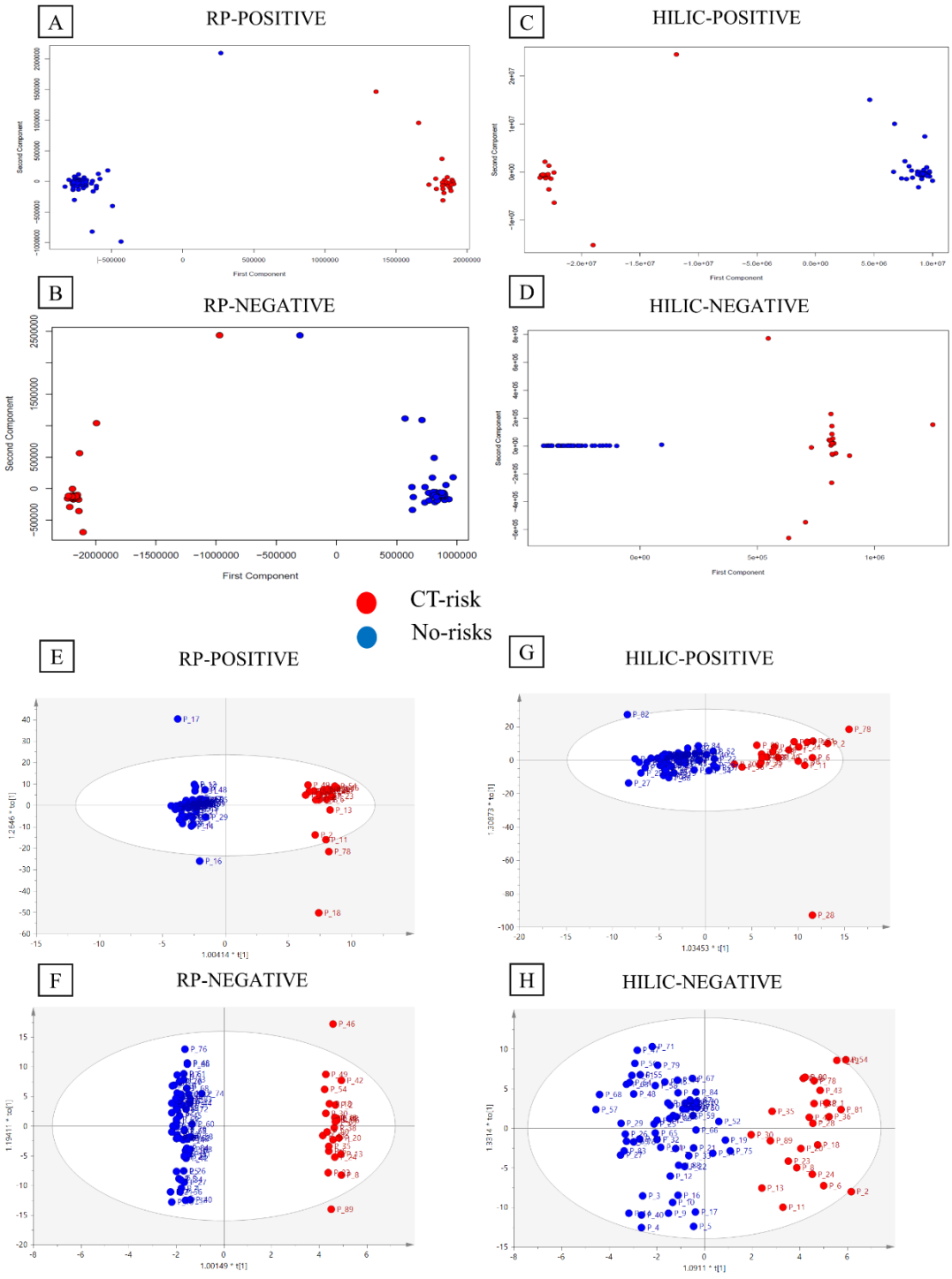

Figure 30 Graphical representation of KODAMA and OPLS-DA classification models: (A) KODAMA topology representation of RP-POSITIVE data; (B) KODAMA topology representation of RP-NEGATIVE data; (C) KODAMA topology representation of HILIC-POSITIVE data; (D) KODAMA topology representation of HILIC-NEGATIVE data; (E) OPLS-DA scores plot of RP-POSITIVE data; (F) OPLS-DA scores plot of RP-NEGATIVE data; (G) OPLS-DA scores plot of HILIC-POSITIVE data; (H) OPLS-DA scores plot of HILIC-NEGATIVE data

3.2.2. Supervised classification

OPLS-DA was chosen to enhance the correlation of specific variables with CT risk, disregarding the orthogonal variation imposed by non-relevant to CT confounders (sex, cancer, age, etc.). The OPLS-DA confirmed the KODAMA observations and provided efficient classification models and many variables with VIP values> 1.5. Both ionization modes of RP analysis, showed better results i.e., decreased intra-group variance and increased inter-groups variance, in contrary to HILIC. Specifically, the scores-plot of RP-POS data, Figure 30, offered the higher degree of separation. These observations suggested that the CT-trends are mainly reflected on non-polar metabolites. The OPLS-DA models were evaluated regarding the predictive accuracy, Supplementary Materials §3.5.

6.3.2.3. BORUTA & FDR t-test

BORUTA algorithm belongs to the family of the “all relevant” variable selection methodologies, and estimates the prediction accuracy of all variables by comparing their performance to the performance of variables with dummy values or the so-called shadow variables, resulting in confirmed, tentative, and rejected variables. The BORUTA algorithm provided 46, 7, 6, 5 confirmed and tentative variables for RP-POS, RP-NEG, HILIC-POS, and HILIC-NEG respectively. In addition, the FDR t-test 124, 46, 224, 24 variables with p-value< 0.05 for RP-pos, RP-neg, HILIC-pos and HILIC-neg respectively.

6.3.3. DIA identification and empirical levels of confidence

The definitive variables (45) were submitted to the identification procedure. In the study, the samples were analyzed using the DIA methodology, which is more information rich concerning the fragmentation data, but lacks in the selectivity of MS2 spectra as implemented in DDA. Therefore, we established an empirical levels of confidence for the DIA MS identification:

Level 5: Definitive features (mz_rt)

Level 4: Molecular formula (using adducts and, isotopes)

Level 3: Metabolite (comparing the experimental MSn with in-silico MS/MS)

Level 2: Comparison of precursor-fragment chromatography

Level 1: Validation using reference standards.

This effort was based on confidence scale regarding the untargeted metabolomics identification, applied to DDA, that was proposed by Schrimpe's et al.[88]. The identified metabolites, along with their statistical scores and their level of confidence are provided in Table 10. More information regarding the identification data are provided in the Supplementary Materials §3.6.

Table 10 summary of CT-risk definitive variables.

mz_rt	KODAMA (Kruskal Wallis lop(*P*.value)	OPLS-DA (VIP)	t-test (*P*.value)	BORUTA	Metabolite name	Level of Identification	ROC (AUC value)	CT-risk regulation
178.0942_2.57	**3.56**	**1.26**	**0.007**	Rejected	5-Hydroxytryptophol	LEVEL 3	0.70	↓
222.113_14.32	**5.66**	**1.34**	**0.001**	Rejected	unknown feature	LEVEL 5	0.73	↓
276.2607_6.04	**5.66**	2.00	**0.003**	**Tentative**	C10H29N9	LEVEL 4	0.50	↓
292.2426_1.43	**4.95**	1.33	**0.000**	Rejected	unknown feature	LEVEL 5	0.74	↓
442.0711_6.31	**4.68**	1.65	**0.001**	**Tentative**	Propinol adenylate	LEVEL 3	0.72	↓
500.8881_1.1	**3.88**	1.35	**0.002**	Rejected	unknown feature	LEVEL 5	0.71	↓
527.4024_5.19	**4.77**	1.57	**0.001**	**Confirmed**	unknown feature	LEVEL 5	0.72	↓
540.3621_5.82	**4.16**	1.34	**0.009**	Rejected	LysoPC(0:0/18:0)	LEVEL 2	0.72	↓
555.3613_3.49	**3.22**	1.05	**0.041**	Rejected	MG(0:0/16:0/0:0)	LEVEL 3	0.56	↓
639.1813_3.14	2.62	**1.77**	**0.002**	**Tentative**	C43H50N4O	LEVEL 4	0.68	↑
112.0449_0.99	**3.42**	**1.60**	**0.004**	Rejected	unknown feature	LEVEL 5	0.70	↑
117.0492_2.35	**3.24**	**1.33**	**0.016**	Rejected	unknown feature	LEVEL 5	0.64	↑
118.0414_0.83	0.68	**1.80**	**0.001**	**Confirmed**	L-Threonine	LEVEL 3	0.71	↑
144.9543_8.25	**3.17**	**1.63**	**0.004**	Rejected	unknown feature	LEVEL 5	0.67	↑
146.9578_8.45	**3.38**	**1.32**	**0.023**	**Tentative**	unknown feature	LEVEL 5	0.64	↑
148.953_8.36	1.77	**1.39**	**0.013**	Rejected	unknown feature	LEVEL 5	0.65	↑
190.9389_8.24	1.82	**1.60**	**0.007**	**Tentative**	unknown feature	LEVEL 5	0.67	↑
291.0676_0.69	**3.46**	**1.21**	**0.033**	**Confirmed**	5-Hydroxyindoleacetylglycine	LEVEL 2	0.66	↑
336.0335_1.33	**3.17**	**1.50**	**0.011**	Rejected	Nicotinic acid mononucleotide	LEVEL 2	0.77	↓
400.8732_8.23	2.03	**1.69**	**0.004**	**Tentative**	unknown feature	LEVEL 5	0.70	↑
404.8688_8.23	**3.19**	**1.45**	**0.012**	**Tentative**	unknown feature	LEVEL 5	0.66	↑
416.8501_8.37	1.55	**1.34**	**0.038**	**Confirmed**	unknown feature	LEVEL 5	0.62	↑
117.0721_0.85	**1.32**	**1.20**	**0.043731**	**Confirmed**	Dihydrouracil	LEVEL 2	0.61	↑
132.0746_0.85	**4.57**	**1.37**	0.005991	Rejected	Creatine	LEVEL 3	0.71	↓

133.0722_0.89	**3.53**	**1.70**	2.08E-07	**Tentative**	2-Hydroxy-3-methylpentanoic acid	LEVEL 2	**0.86**	↓
138.9059_0.77	**3.10**	**1.52**	**0.027**	Rejected	unknown feature	LEVEL 5	0.65	↓
147.0746_0.81	**3.01**	**1.56**	0.087	**Tentative**	unknown feature	LEVEL 5	0.46	↑
163.1302_9.58	**3.53**	**1.58**	**0.014**	**Tentative**	Hydroxyoctanoic acid	LEVEL 2	0.66	↑
166.0813_2.78	**3.00**	**1.70**	1.12E-04	Rejected	unknown feature	LEVEL 5	**0.83**	↑
170.055_1.6	2.18	**1.54**	**0.030**	**Tentative**	L-Isoleucine	LEVEL 3	0.64	↑
172.1305_6.22	**4.44**	**1.43**	**0.004**	**Confirmed**	4-Hydroxynonenal	LEVEL 2	0.68	↑
182.0628_5.87	**3.30**	**1.51**	**0.032**	Rejected	2-Hydroxyphenethylamine	LEVEL 3	0.66	↑
188.068_3.98	**3.09**	**1.51**	5.59E-08	**Confirmed**	unknown feature	LEVEL 5	**0.85**	↓
203.0505_0.92	**3.77**	**1.32**	**0.040**	Rejected	Formiminoglutamic acid	LEVEL 2	0.60	↑
211.0934_10.16	**4.75**	**1.52**	**0.012**	**Tentative**	Nonate	LEVEL 2	0.69	↑
236.1107_7.9	**3.57**	**1.52**	**2.13E-10**	Rejected	Glutamylalanine	LEVEL 2	**0.88**	↑
256.941_0.71	**3.30**	1.08	**0.031**	Rejected	unknown feature	LEVEL 5	0.64	↓
261.0087_9.21	**3.57**	1.17	**0.038**	**Confirmed**	5-Methylthioribose 1-phosphate	LEVEL 2	0.70	↓
263.0848_13.5	**4.35**	**1.60**	**0.002**	Rejected	unknown feature	LEVEL 5	0.71	↓
283.7151_9.63	0.78	**1.32**	**0.036**	**Tentative**	unknown feature	LEVEL 5	0.61	↑
297.6866_8.3	2.80	1.01	**0.002**	**Tentative**	unknown feature	LEVEL 5	0.70	↑
374.2361_5.7	2.30	**1.42**	**0.015**	**Confirmed**	Dodecanedioylcarnitine	LEVEL 2	0.70	↑
384.2934_12.94	2.12	**1.55**	**0.032**	Rejected	(R)-3-Hydroxy-hexadecanoic acid	LEVEL 2	0.65	↓
576.3583_10.92	**3.00**	1.10	4.60E-16	**Confirmed**	3-Hydroxy-9-hexadecenoylcarnitine	LEVEL 2	**0.97**	↑

K.Wallis log(*P* value) is the value of Kruskal Wallis t-test, resulting from the KODAMA analysis; The VIP values were obtained by the Ortho Partial Least Square -Discriminant Analysis (OPLS-DA); the t-test (*P* value) resulted from the univariate t-test and; the AUC value was provided by receiver operating characteristic (ROC) analysis; BORUTA shows the confirmed/ tentative/ rejected variables, resulting by the respective algorithm; ; the last column includes information regarding the level of identification confidence, i.e., Level 5 (definitive variable); Level 4 (Formula Prediction); Level 3 (Metabolite Prediction); Level 2 (Metabolites' confirmation using high-collision energy mass spectrometry (MS) data. The values in **bold** are those passed the statistical cutoffs.

6.4. Discussion

The major goal of this study was to provide reliable metabolomics evidence regarding the early assessment of the CT risk. The statistical methodology focused on the naive hypothesis for the existence of a CT-risk metabolic fingerprint using. Therefore, the metabolic profiles of the patients (before the chemotherapy, day=0), were grouped solely with the *a posteriori* knowledge of CT expression. Considering the fuzziness introduced by cofactors (sex, age, and cancer) on children's metabolic profiles, it was deemed essential to develop a strategy able to increase the confidence of the discriminating variables. Therefore, in contrast to the common metabolomics approaches (routinely using OPLS-DA and/or ROC analysis for biomarker detection), here we combined four chemometric variable-selection approaches whose main theoretical foundations can be regarded as complementary: i) KODAMA is based on the creation and the cross validation of random classification models; ii) OPLS-DA belongs to the class of data-reduction methodologies; iii) BORUTA is an "all-relevant" variable selection method, that assesses the statistical significance of all the variables, by comparing them with dummy values, and at last; iv) FDR-t.test is the most commonly used univariate tool. Based on the Bayesian probabilistic theory, the variables provided as discriminative by more than one methods, termed definitive variables, exhibited an increased probability of being CT-risk-related. Moreover, aiming to increase the identification reliability of the resulted DIA data, we established an empirical scale of identification confidence, modifying the respective scale of Schrimpe et al.[88], applied on untargeted DDA identification.

Summarizing the results, the research confirmed the existence of early metabolomics evidence, detectable before the start of the chemotherapy, betraying the risk of future CT expression. The majority of the identified metabolites are hydroxy derivatives of amino acids, fatty acids and lipids. The altered levels of hydroxy fatty acids (HFAs) and hydroxy carnitines suggest beta-oxidation dysregulation, perhaps leading to oxidative products, and mitochondrial dysfunction[89],[90]. Furthermore, the elevated levels of long-chain HFAs indicate dysregulation of FA mitochondrial beta-oxidation and; increase of peroxisomal oxidation resulting in formation of oxidative products,

as H_2O_2 [91]. The increase of small-chain HO-FAs in CT-Risk patients, i.e., the hydroxyoctanoic acid, could indicate a parallel occurring dysregulation of FA metabolisms, i.e., elevation of peroxisomal β-oxidation of long-chain FA, resulting in smaller-chain in FA[34]. The alterations of medium-chain 3-hydroxy-FA is also associated to Fatty Acids oxidation disorders (FAOD), inborn errors of metabolism resulted by dysregulation of mitochondrial β-oxidation[92]. The alterations of 4-HNE combined with the dysregulated lipids content indicate alterations on lipids peroxidation[19-21], which is also linked to FA-metabolism dysregulation as well[35]. *Beuchel et.al.*, recently highlighted associations of mitochondrial and peroxisome FA oxidation along with acylcarnitines dysregulation with the decline of cardiovascular health[93]. Moreover, the differentiations of 2-hydroxy-3-methylpentanoic acid[94], formiminoglutamic acid[95], and glutamylalanine (AUC=0.88) are implying also amino acids metabolism dysregulation. In parallel the amino acids pathways may be indirecltly disturb by the decline of 5-methylthioribose 1-phosphate that impairs the biosynthesis of methionine [96] .

6.4.1. 3-Hydroxy-9-hexadecenoylcarnitine

3-hydroxy-9-hexadecanoylcarnitine, belongs to the class of long-chain acylcarnitines and can be considered as potential prognostic biomarkers, as showed AUC= 0.97 and reached the 2nd level of identification confidence. The compound was increased in CT-risk patients. Acylcarnitines are involved in the transportation of organic and fatty acids from the cytoplasm to mitochondria, thus, their perturbed levels imply dysregulation of fatty-acids metabolism. Corresponding observations of this metabolite's upregulation in blood have been reported in systolic and diastolic heart failure and also in chronic heart failure [97],[89],[98].

6.4.2. 4-Hydroxynonenal (4-HNE)

4-Hydroxynonenal, 4-HNE, was found increased in CT-risk group, with AUC = 0.68. Despite its moderate prediction ability, the altered levels of this metabolite constitute an interesting observation. Previous studies have showed connection between 4-HNE and CT-occurrence, as it was increased in events of *doxorubicin (DOX)* induced CT. In DOX-induced-CD, 4-HNE and HNE-adducts seem to participate in the mechanisms of cell-apoptosis and mitochondrial

dysfunction[99],[100]. 4-HNE is a normal endogenous metabolite, formed as a secondary product of lipids peroxidation. It is increased in tissues at redox conditions and thus, it is considered as biomarkers of oxidative stress as well. Due to its chemical structure is highly reactive molecule and forms several protein adducts leading to cell-signaling-alterations, negatively affecting the life and the death of cells. Also, its upregulation has been linked to cardiovascular, neurodegenerative and metabolic diseases [101],[102],[103].

6.4.3. Nicotinic acid mononucleotide

The observed decrease of nicotinic acid mononucleotide (NMN, AUC=0.77) levels is also remarkable, as it participates in cofactor biosynthesis and in metabolic pathways for the synthesis of nicotinate and nicotinamide. NMN is a substrate for nicotinamide riboside kinase in the biosynthesis of NAD[104]. The dysregulation of NMN also confirms the dysregulation of tryptophan metabolism. NAD+ participates in organism's redox mechanisms, and its decline has been associated to normal aging [105]. In the current case, the low-levels of NMN could indicate inhibition of redox procedures in the CT-risk patients.

Summarising the above, the evidence of the investigation suggest that CT-risk patients express early abnormalities in the metabolic pathways that are directly/indirectly linked to amino-acids, fatty acids, and lipids regulation. These pathways are involved in the general pathway of cardiac energy metabolism [93]. It is worth mentioning that the examined cohort showed no clinical evidence of cardiac dysfunction before the start of the chemotherapy. The above indicates that the early, subclinical, alterations of the CT-risk patient's metabolism, were further triggered by the anti-neoplastic agents, leading to clinical evidence of CT expression.

6.5. Materials and Methods

6.5.1. Examined Cohort

The study has been approved by the Scientific Council of the Children's Hospital "AGIA SOFIA", Athens, Greece (Protocol No.: 27148/23.12.2019); and the Scientific Council of the Hippocrate Hospital, Athens, Greece (Protocol No.: 20162/11/12/2018); and has also received the informed consent of the children's parents or legal guardians. Blood samples of 89 oncology pediatric

patients were obtained by the medical personnel of the Department of Pediatric Hematology-Oncology (T.A.O.) of the "AGIA SOFIA" Children's Hospital. The sampling preceded any administration of antineoplastic drugs. The blood samples were instantly centrifuged, and plasma was collected and stored at -80^{o}C until the start of the metabolomics analysis. After blood sampling all children were submitted to the proper chemotherapy protocol and 26 of them expressed an early or an acute CT event, in the course of their treatment. These events were diagnosed by clinical evidence, i.e., imaging techniques and laboratory findings. The *a posteriori* knowledge of CT was used to group the examined cohort into CT-risk (26) and No-risk groups.

6.5.2. Sample preparation and Data acquisition

600 μL of frozen methanol were added mixed with 200 μL of plasma and vortexed for 20 s. The diluted samples were centrifuged using a NEYA 16R centrifuge (REMI, Mumbai, India) at 10000 rpm, 5 min, 4^{o}C. Aliquots of 200 μL supernatant were evaporated until dryness by a HyperVAC-LITE centrifugal vacuum concentration (Hanil Scientific Inc, Gimpo, Korea), the dried samples were stored at -80^{o}C and reconstituted before the analysis with 50 μL of IS standard solution[106],[107],[108], as it is described in the Supplementary Materials. Pooled samples were analyzed together with the samples as Quality Control (QC) samples. For the chromatographic analysis both HILIC and RP were employed, as complementary methodologies for the separation of polar and non-polar metabolites respectively. A Thermo Fisher Scientific *Acclaim RLSC C18 column* (2.1x100 mm, 2.2 μm) and an *ACQUITY* Waters *UPLC BEH Amide* column (2.1x100 mm, 1.7 μm) were used respectively for the RP and HILIC separation. The Bruker *Maxis* Impact QTOFMS (Bruker Daltonics, Bremen, German) coupled to a Thermo *Dionex Ultimate 3000 LC* system (Thermo Fischer Scientific, Dreieich, Germany) was employed for data acquisition, using ESI ion-source. The broad-band Collision Induced Dissociation (bbCID) mode was applied to obtain low-collision energy (CE) and high-CE MS data. More information are provided in the Supplementary Materials §3.3.

6.5.3. Peak-picking & data pre-processing

MZmine 2.51 [109] was employed, and peak-picking procedure was based on our previous protocol for DIA pre-processing[52]. The low-CE data were submitted to the peak-picking procedure, aiming towards the biomarkers' determination. The high-CE information was used during the final step, to facilitate the metabolites identification. The feature-lists were submitted to retention-time-drift correction, employing the "well-behaved" peaks. The RANSAC algorithm was used for the alignment. The QCRFSC [76] algorithm was employed for QC based signal-correction. The NA.filter was set at 0.8 in order to include only the variables with non-zero values for the 80% of the samples. The %CV cutoff was set at 40 to exclude the variables with relative standard deviation (RSD) higher than 40%. Afterwards, yohimbine and reserpine compounds (ISs) were used for IS-based signal correction, employing the *Normalization using optimal selection of multiple internal standards* (NOMIS) algorithm[110] playing in NOREVA2.0 platform [111]. Then, the data were transformed (LogTranformation), and scaled (Unite Variance scaling) in order to ensure their normality. PCA was used to investigate and exclude the outliers. As outliers were considered the samples that were out of the Hotteling's T2 (95%) ellipse.

6.5.4. Statistical methods & tools

For the statistical interpretation of the data the study used the Knowledge discovery by accuracy maximization, KODAMA, algorithm [47] and the BORUTA [46] algorithm, as they implemented in the R statistical language environment; and also, the OPLS-DA supervised classification methodology, conducted using SIMCA 14.1 (MKS Umetrics, Sweden). Also, the false discovery rate (FDR) t.test and the receiver operating curve (ROC) analysis were conducted using the MetaboAnalyst 5.0 online platform.

6.5.4.1. Definitive variables

In order to reduce the amount of the variables, based on their statistical importance, specific thresholds were set for each statistical methodology. Thus, the threshold for KODAMA, OPLS-DA and FDR-t.test were Kruskal Wallis p.value >3, VIP value >1.5 and p.value <0.05 respectively. Regarding BORUTA, the *confirmed* and the *tentative* variables were considered as important. As

definitive variables were considered those who passed the thresholds of three or more statistical trials.

6.5.5. Identification

The ion chromatograms of the definitive variables were extracted and manually evaluated (S/N, reproducibility) using Data Analysis software (Bruker Daltonics, Bremen, Germany). Both low and high collision energy (CE) data were used for the identification. The low-CE MS were used for the molecular formula prediction using the isotopic profile, and the high-CE MS were compared to in-silico data to identify specific metabolite. MyCompoundID online library was used for the identification procedure, using the MS-fragmentation data provided by the HMDB. Specifically, the XIC (mz_rt) of the definitive variables were ion-extracted in the raw spectra with mass and retention time tolerance of 5 mDa and 0.05 min, respectively. Low-CE MS were used for the molecular formula determination using the SmartFormula algorithm of DataAnalysis software (Bruker Daltonics, Bremen, Germany). The mSigma threshold was set to 20. In cases that SmartFormula provided no potential formula of acceptable score, the MS-search module of myCompoundID (MCID) [37] online library was employed. In the latter case, the experimental isotopic ratio was manually calculated and compared to the reference isotopic ratio, as provided by Compass IsotopePattern software (Bruker Daltonics, Germany). Then, the high-CE MS data were assigned to the corresponding precursors via retention time (after removing the background signal). The cleaned high-CE-MS clusters were considered as pseudo-MSMS, as DIA selected no precursor ion, and were employed for the verification of compounds structure. The pseudo-MSMS (mz_rt and %I) were uploaded on MS/MS-Search module of MCID and compared to the *in-silico* predicted MSMS provided by the online library. The metabolites that were identified using the in-silico fragmentation were submitted to further verification, to validate that those fragments belong to the precursor of interest, as DIA does not assign the fragments to a specific precursors. Thus, the XICs of the 3 most abundant theoretical fragments as provided by HMDB (Predicted/Experimental MS/MS spectra), were extracted in the raw data, setting 10 mDa as the mass accuracy threshold; and their chromatographic

shape was empirically compared to the XIC of the parent ion, as described in Figure 31.

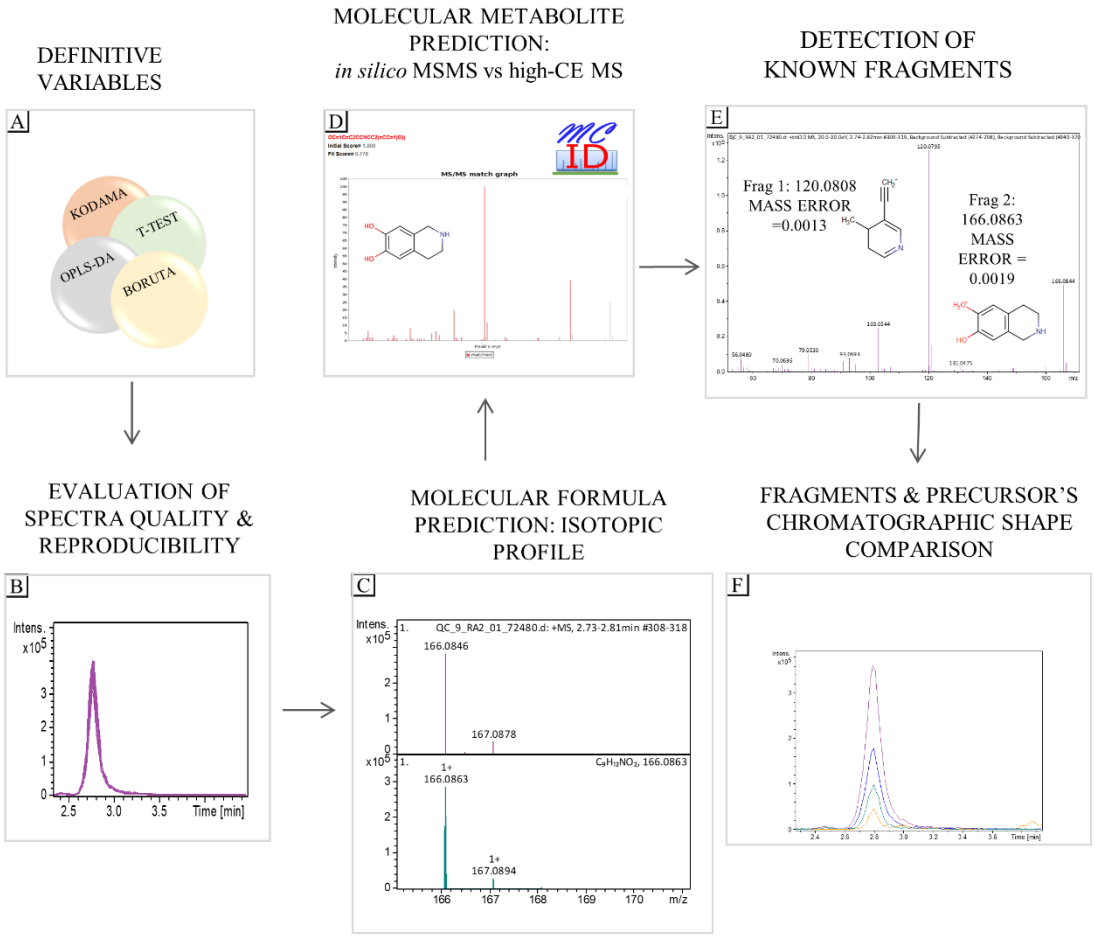

Figure 31 Graphical representation of the identification pipeline: (A) The variables that pass the cutoffs of three or more statistical tests (KODAMA, OPLS-DA, BORUTA, FDR-t.test) are considered as definitive variables; (B) The mz_rt of the definitive variables are extracted as XICs from the raw data and evaluated (S/N and reproducibility); (C) The isotopic profile of the definitive variables is compared with reference ones, during the molecular formula determination; (D) The molecular formula is confirmed by the comparison of the experimental high-CE MS with in-silico MS, and was attributed to specific metabolite; (E) The three more abundant theoretical fragments of the metabolites are extracted as XICs in the hight-CE MS data and their mass error is evaluated; (F) The peak shape of the precursor and the fragments are empirically compared.

6.6. Conclusion

Cardiotoxicity is a common side effect of anti-neoplastic treatment protocols and is expressed in both adults and children patients, leading to increased morbidity and mortality. However, the ratio of the chemotherapy induced CT expression indicates that the condition is related to the individual's phenotype and biochemical background. Thus, the investigation for early, subclinical evidence related to future CT-expression could simplify the risk assessment and the early prevention of this condition. Moreover, it would facilitate the application of tailored therapeutic schemes. Herein, the current study examined a cohort of children diagnosed with malignancies and solid tumors, aiming to estimate the risk to display CT when they undergo chemotherapy. A metabolomics analysis protocol was applied to plasma samples that were obtained before the start of chemotherapy, and the experimental observations were combined with the *a posteriori* knowledge of CT expression in a post-hoc analysis. The final results provided novel evidence demonstrating the existence of early detectable metabolic markers related to the risk of CT occurrence. All multivariate models classified the CT-risk children with increased reliability. The CT-risk children exhibited alterations in pathways linked to cardiac energy metabolism, suggesting that chemotherapy is not the sole reason for CT expression. Probably, chemotherapy triggers preexisting cardiac function abnormalities, accelerating the establishment of cardiac dysfunction incidents.

7. Chapter 7

Metabolomics investigation for the impact of the human doses of colistin in the kidney and liver of mice

7.1. Abstract

Colistin (CMS) is used for the curation of infections caused by multidrug-resistant bacteria. Its use is constrained by toxicity, particularly in kidney and neuronal cells. The recommended human doses of CMS are 2.5–5 mg/kg/day, and the toxicity is linked to higher doses or long-term administration. So far, the in vivo toxicity studies have used doses even 10-fold higher than human doses. It is deemed essential to investigate the impact and the metabolic response for the CMS-human doses, where there is no clinical evidence of toxicity. Therefore, in the current study, two doses of CMS, Low (1 mg/kg) and High (1.5 mg/kg) versus a control (normal saline), were administered to mice, and samples of plasma, kidney, and liver were analyzed. A thorough metabolomics workflow combined with univariate and multivariate statistical interpretation was performed. The results pointed out six dose-responding metabolites (PAA, DA4S, 2,8-DHA, etc.), renal dopamine dysregulation, and extended perturbations in renal purine metabolism. Also, the study determined altered levels of liver suberyglycine, a metabolite linked to hepatic steatosis. An intriguing finding was the increased formation of renal xanthine, which is an AChE activator, leading to rapid degradation of achetylcholine, suggesting an association of nephrotoxicity and neurotoxicity.

7.2. Introduction

Colistin is typically used as a last resort antibiotic to treat infections caused by multidrug-resistant bacteria as it has a narrow therapeutic index and can cause significant toxicity, particularly to the kidneys and nervous system [112]–[114]. It belongs to the class of antibiotics known as polymyxins, which have a unique mechanism of action that disrupts the bacterial cell membrane. This mechanism of action makes polymyxins effective against many Gram-negative bacteria that are resistant to other antibiotics bacteria such as *Pseudomonas aeruginosa, Acinetobacter baumannii*, and *Klebsiella pneumoniae* [115]. However, polymyxins are also toxic to human cells, particularly kidney cells and neuronal cells. The drug is usually given as colistimethate sodium (CMS), prodrug of

colistin[116]. The dosage regimen of CMS can vary depending on the patient's age, weight, the severity of the infection, and the method of administration (injection or inhalation). For the treatment of infections caused by Gram-negative bacteria, the usual adult dose is 3 million units (MU) daily, administered intravenously in divided doses every 8 hours[117]. An average steady-state plasma colistin concentration of 2 mg/L seems to be a reasonable target value [118]. A loading dose CMS is often recommended to achieve a therapeutic concentration quickly. The loading dose can vary depending on the patient's condition, but it is usually higher than the subsequent maintenance dose [119], [120]. The duration of treatment can also vary depending on the severity of the infection and the patient's response to therapy. According to the prescribing information for CMS injection, the maximum daily dose should not exceed 5 mg/kg of body weight (https://www.accessdata.fda.gov/drugsatfda_docs/label/2009/050108s026lbl.pdf, accessed on 26 April 2023). It is important to note that the dosage regimen for CMS should be adjusted in patients with renal impairment, as the drug is primarily eliminated by the kidneys [121]. The use of CMS should always be monitored by a healthcare professional.

CMS, like other antibiotics, can cause side effects and toxicities, i.e., brain dysfunction and neurotoxicity. The mechanism of CMS-induced neurotoxicity is not fully understood, but it is thought to involve the drug's ability to penetrate the blood-brain barrier and interact with neuronal cells [122]–[124]. Furthermore, CMS can cause damage to the kidneys, especially if it is used for a long time or at high doses. This can lead to symptoms such as decreased urine output, swelling in the legs or feet, and shortness of breath[125]–[127]. Although CMS is primarily associated with kidney toxicity, there is some evidence to suggest that it may also have an effect on the liver[128]. It is important to note that the risk of toxicities may vary depending on the dose, duration of treatment, and individual patient factors. Therefore, the levels of CMS should be carefully monitored.

The aim of this study is to shed light on the mechanisms underlying the metabolic alterations induced by the modest dosed of CMS administration. For this reason, an in-vivo stimulation of CSM metabolic changes was performed,

administrating low doses of CSM (1 and 1.5 mg/kg) in mice, aiming to identify and metabolites related to CSM and to understand how their levels change in response to CSM-dose. Until now, one metabolomics study has been conducted for the characterization of urinary metabolites as biomarkers of colistin-induced nephrotoxicity in rats[129]. Furthermore, Long et al.[128] have studied the alteration of kidney and liver metabolome after administration of high CMS dose. Herein, a metabolomic analysis has been performed in kidney, liver, and plasma after administration of human-doses of CMS in mice.

7.3. Statistical analysis results

The Control (C), the Low-dose (LD), and the High-dose (HD) treated mice were compared in pairs to detect classification trends correlated i) to the drug administration (C vs LD) and, ii) to the drug dose (LD vs HD). Initially PCA exhibited clear separation of the C and LD mice only in the kidneys, Figure 32A. The OPLS-DA resulted in effective classification models for all abovementioned comparisons, for all biosample types (plasma, kidney, liver), Figure 32(G-L). The confidence and the performance of the developed models were estimated through miss-classification and permutations testing and model ROC analysis, and the results.

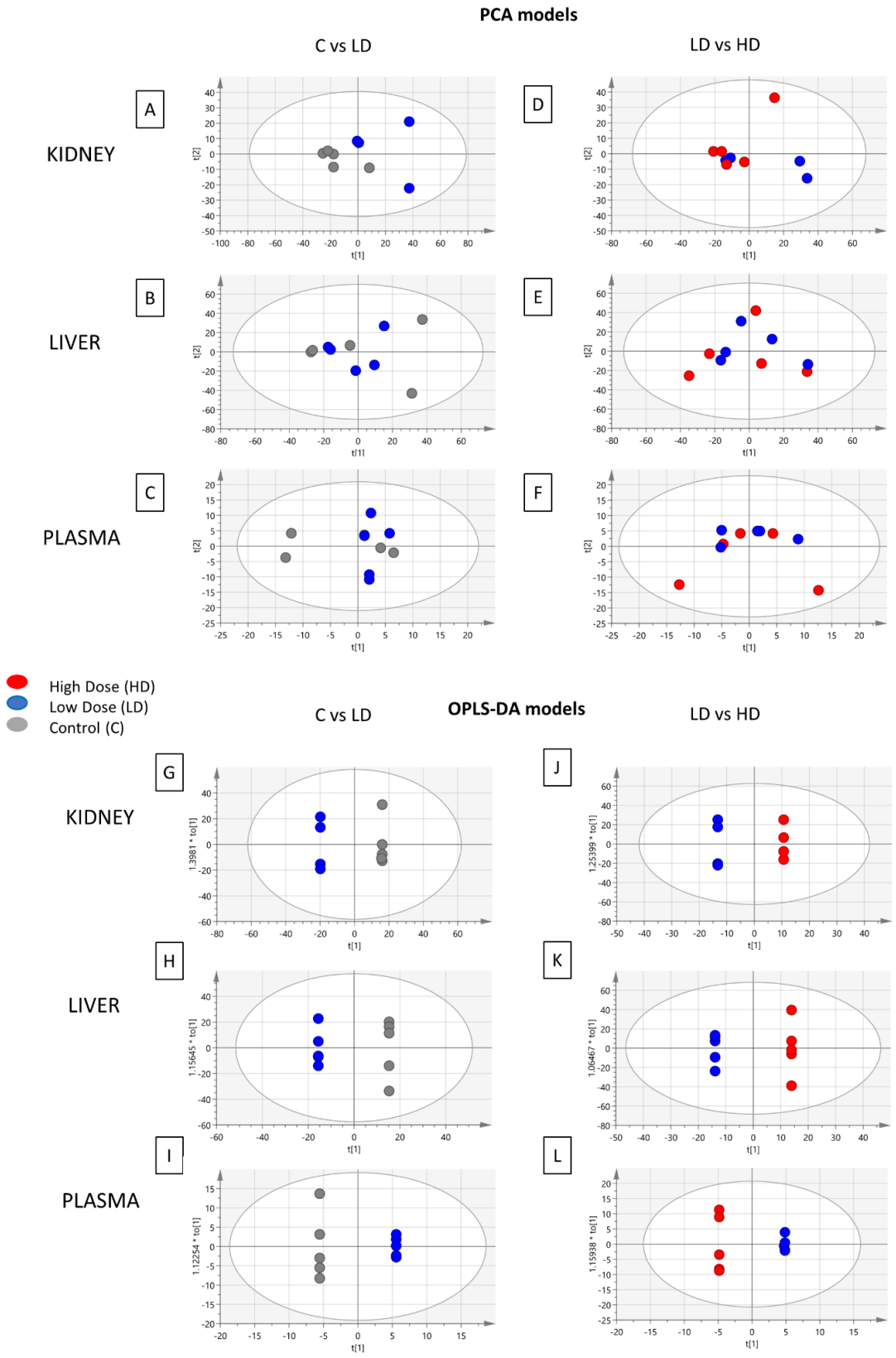

Figure 32 Summary of the PCA and OPLS-DA results of the ESI+ datasets. PCA scores' plots of: (A) Kidney, C-LD; (B) Liver, C-LD; (C) Plasma, C-LD; (D) Kidney, LD-HD; (E) Liver, LD-HD; (F) Plasma, LD-HD. OPLS-DA scores' plots: (G) Kidney, C-LD; (H) Liver, C-LD; (I) Plasma, C-LD; (J) Kidney, LD-HD; (K) Liver, LD-HD; (L) Plasma, LD-HD.

The study focused on the alterations provoked by the increase of the dose, thus, the differentiated variables from LD-HD comparisons with VIP> 1.2 were employed for univariate ROC curve analysis in order to detect the variables that response to the dose increase. The number of variables with AUC value > 0.8 were, 251 variables for kidneys (49 up-regulated in the HD), liver 345 variables for liver (136 up-regulated in the HD), and 47 variables for plasma (28 up-regulated in the HD).

Moreover, PLS models were employed to investigate the linear correlation between the three levels of administered CMS (0, 1 and 1.5 mg/kg) and the alteration occurring to the mice's metabolomic profiles. The excellent linearity (R^2> 0.99) and the low root means square error of estimation (RMSEE) of the PLS models, Figure 33, prove that there is a linear correlation between the administered dose and the resulting metabolomics alterations.

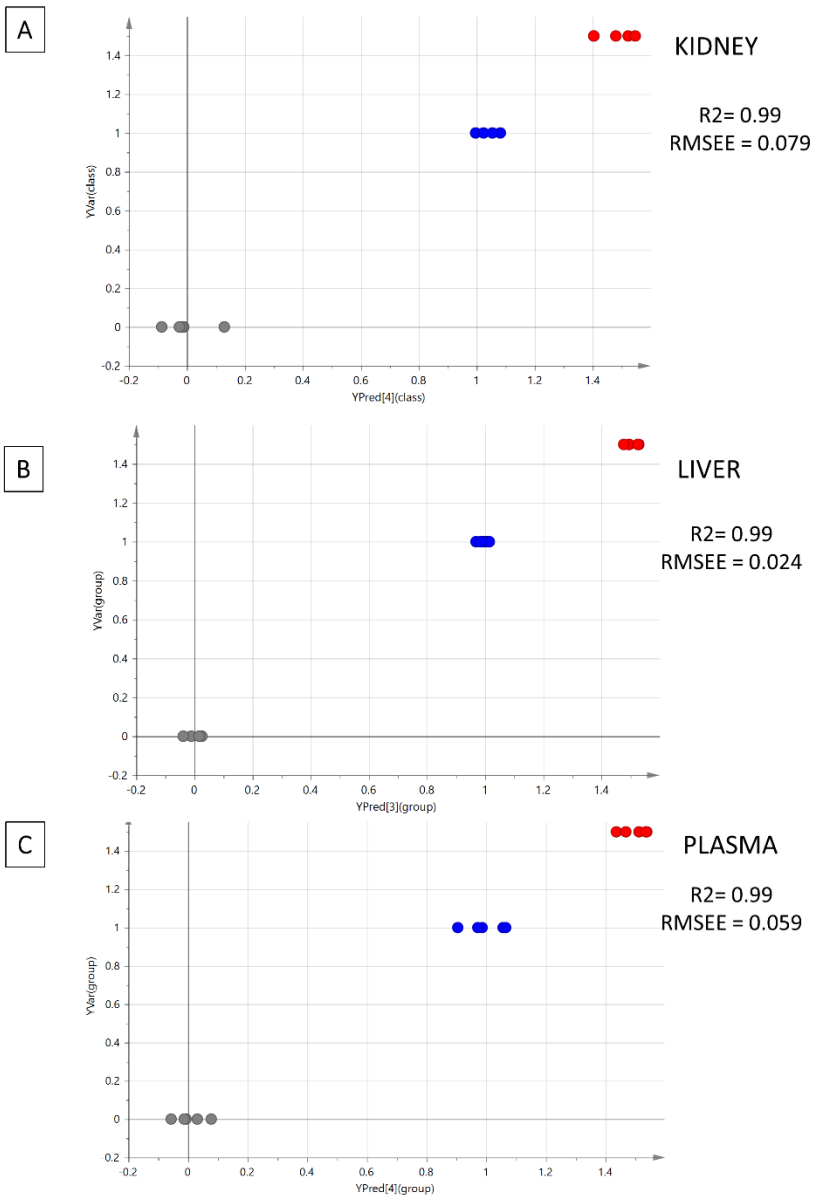

Figure 33 Summary of PLS analysis. Observed vs predicted plots of ESI+ datasets: (A) Kidney; (B) Liver; (C) Plasma. The grey spots represent the C samples ((CMS 0 mg/kg); the blue spots represent the LD-samples (CMS 1.0 mg/kg); and the red spots represent the HD.

To elaborate on this observation, the OPLS-DA models of C-LD and LD-HD comparisons were used to build shared and unique structure plots, known as SUS-plots, aiming to highlight the most highly dose-correlated variables. The SUS-plot describes the correlation of the predictive variables afforded from the two OPLS-DA models, by plotting the loadings of both against each other. In this case, the 2 OPLS-DA models share a common group, the LD group. The SUS-plot in Figure 34A represents the loadings of C-LD and LD-HD for the

dataset of kidney, ESI+. The features at the edges of y-axis are discriminant for the model C-LD and those at the edges of x-axis are discriminant for the model LD-HD. The features existing in the diagonal from the lower left to the higher right, (Figure 34A, green arrow), are the those that present shared structures, whereas the existing in the diagonal from higher left to lower right (Figure 34A, black arrow) are presenting unique structures. As the LD is the common group in the two OPLS-DAs, the variables of shared structure show the same regulation in the LD group, in both C-LD and LD-HD comparison. For example, the variable in Figure 34B shows increased levels in LD samples and decreased levels in C and HD samples, so, as it is always increased in LD group is considered as variable of shared structure. On the other hand, the variable in Figure 34C shows increased levels in LD group, comparing to C-group levels, and in parallel shows decreased levels in LD group when it is compared to HD. Thus, it is considered as a variable of unique structure. The features of unique structure seem to express dose-response, while those of shared structure do not provide meaningful information.

Aiming to determine variables showing linear correlation to the dose, a pipeline of three steps was followed: i) at first, the variables existing at the edges of the shared-structure-diagonal (of the SUS-plot) were selected; ii) then, they were submitted to pairwise (C-LD and LD-HD) univariate ROC curve analysis and those having AUC > 0.8 in both comparisons were kept and; iii) they were used to develop regression curves, as shown in Supplementary Figures 4.1 & 4.2. This procedure resulted in 16 dose-response variables, with $R^2 > 0.7$, the 9 of them were upregulated in correlation to the dose, Table 11.

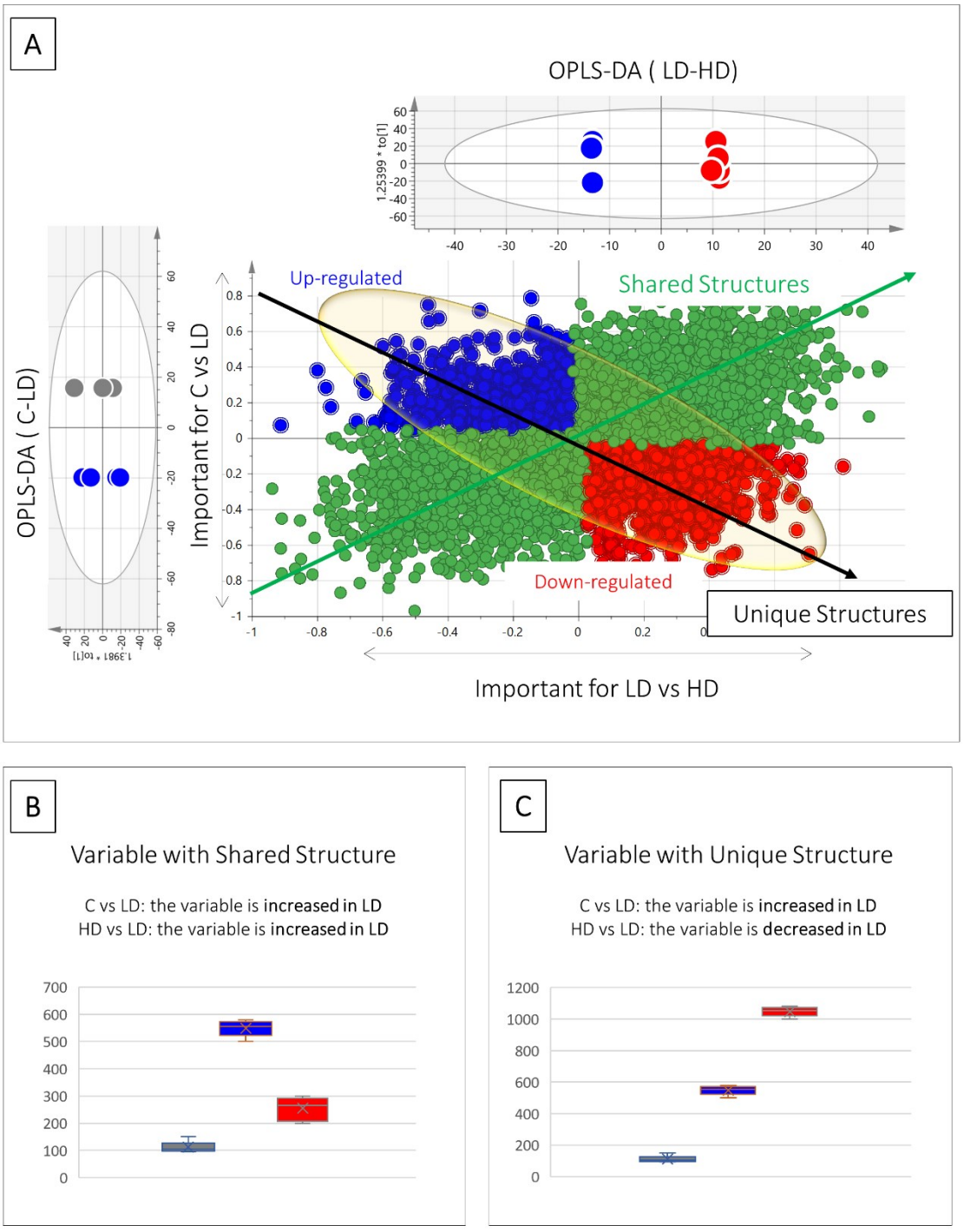

Figure 34 Graphical description of the SUS-plot basic information. (A) SUS-plot developed by the OPLS-DA models of C-LD and LD-HD of kidney ESI+ dataset. The green spotted area represents the variables of shared structure. The blue and the red spotted areas represent the variables of unique structure. At the edges of the black arrow of unique structures exist the statistically significant dose-response variables (red: upregulated, blue: down regulated); (B) A box-plot representing a variable of shared structure, which shows increased levels in LD (blue) but decreased levels in C (gey) and HD (red); (C) Box-plot representing an example of an up-regulated unique structure variable which is increased in LD (blue) compared to C (grey) and in parallel is decreased in LD, compared to HD (red).

7.4. Variable selection and identification

The identification was statistically driven by the results of the pairwise comparison of LD-HD and also, by the results of the SUS-plot procedure. The pipeline followed for the current metabolomics study is described in Figure 35. Thirty-four statistically important features were attributed to known metabolites, as shown in Table 2. The identified metabolites were submitted to pathway analysis, showing extended alteration of purine metabolism in kidneys. Thus, an additional hypothesis driven peak-picking, focusing on the metabolites involved in the purine metabolism pathway, was performed. The screening list of the investigated metabolites is provided in the Supplementary Material, Table S3. For this step the C, LD and HD group of kidney samples were used, and 27 metabolites were finally identified. These metabolites were further submitted to pathway analysis, also employing their peak-area signals (AUC) as additional information to estimate the alterations in purines pathway.

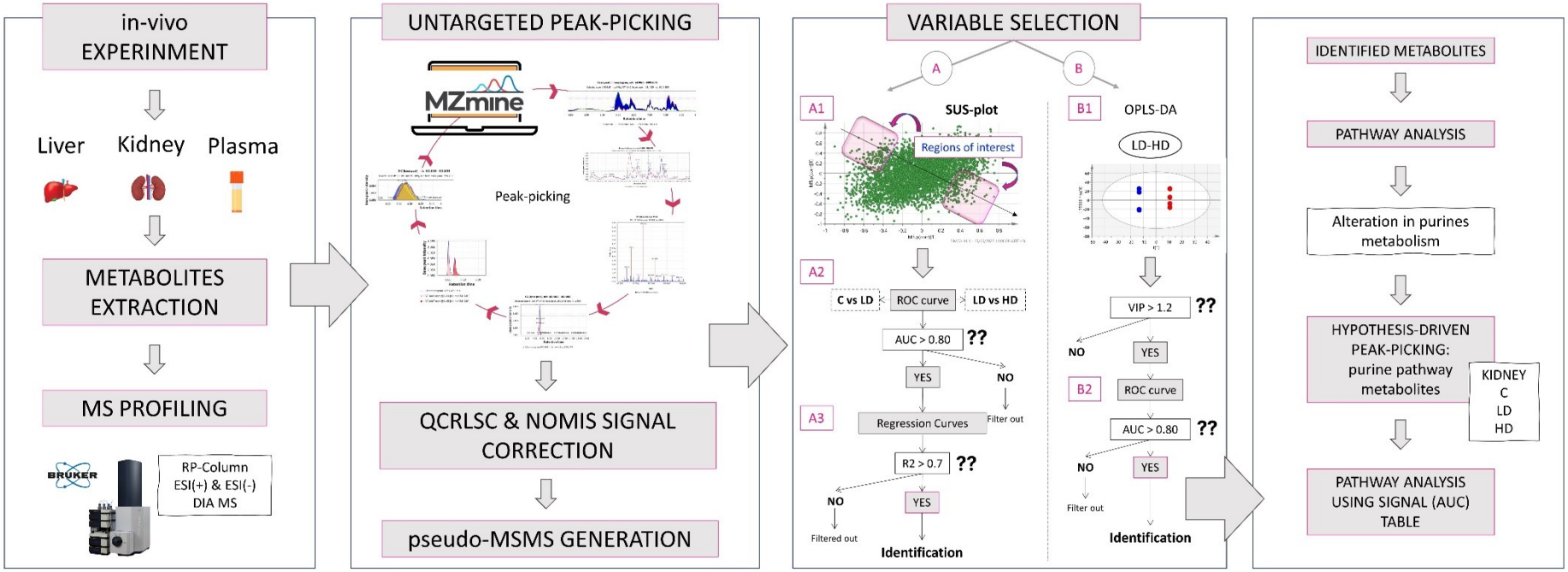

Figure 35 Graphical description of the applied metabolomics workflow: Initially the in-vivo experiment simulate the impact of 1 and 1.5 mg/kg CMS in mice. The liver, kidney and plasma samples were used for metabolites extraction, and the analysis was conducted with a Bruker maxisImpact QTOF MS, using RPLC and ESI+/- ionization. The MS was acquired with DIA methodology. Then the raw data were used for untargeted peak-picking and the feature list was submitted to QC and IS based signal correction. RamclustR was used for the generation of pseudo-MSMS. Then, the features were used for variable selection performed with 2 ways: i) extraction of dose-response variables using SUS-plot procedure and; ii) extraction of dose-correlated variables combining multivariate (OPLS-DA) and univariate (ROC) models of LD-HD comparison. The most discriminant variables were subjected to identification. The identified metabolites were used for a naïve pathway analysis. The results showed alteration in renal purine metabolism and thus a targeted peak-picking of specific metabolites was applied in the raw data of kidneys. The results were used for a new semi-quantitive pathway analysis.

Table 11 List of identified metabolites, resulted from Colistin RPLC untargeted analysis.

mz_rt	*Precursor*	*Common Name*	*Theoretical Mass (Da)*	*Formula*	*Mass Error (Da)*	*Initial Score*	*Fit Score*	*Dataset*	*1reaction*	*AUC (LD-HD)*	*Regulation*
192.0526_4.06	[M-H]-	Glycolic acid	76.016045	C2H4O3	-0.000099	1	0.675	KIDNEY(-)	[+ (C5H5N5 - H2O)]	1	↓
211.9876_3.93	[M-H]-	L-Aspartyl-4-phosphate	213.003842	C4H8NO7P	-0.008966	1	0.383	KIDNEY(-)	-	1	↓
125.0185_3.57	[M+H]+	(R)-N-Methylsalsolinol	193.110279	C11H15NO2	-0.001585	1	0.931	KIDNEY(+)	[+CO2]	0.85	↑ dose-dependent
136.0652_1.76	[M+H]+	Adenine	135.054495	C5H5N5	0.003429	1	0.458	KIDNEY(+)	-	1	↓
159.0317_1.5	[M+Na]+	Hypoxanthine	136.038511	C5H4N4O	0.003965	1	0.494	KIDNEY(+)	-	1	↑
175.006_1.51	[M+K]+	Phenylacetic acid	136.05243	C8H8O2	-0.009592	1	0.302	KIDNEY(+)	-	0.9	↓ dose-dependent
194.0826_5.37	[M+H]+	2-Methylhippuric acid	193.073894	C10H11NO3	0.00143	1	0.707	KIDNEY(+)	-	1	↓
232.0382_5.38	[M+K]+	2-Methylhippuric acid	193.073894	C10H11NO3	0.001144	1	0.627	KIDNEY(+)	-	1	↓
256.122_5.57	[M+Na]+	Hydroxypropionylcarnitine	233.126324	C10H19NO5	0.006452	1	0.321	KIDNEY(+)	-	1	↓
262.062_4.45	[M+H]+	2,8-Dihydroxyadenine	167.044325	C5H5N5O2	-0.006314	1	0.508	KIDNEY(+)	[+C4H2N2O]	1	↑ dose-dependent
269.0943_2.25	[M+H]+	Inosine	268.080771	C10H12N4O5	0.006253	0.998	0.822	KIDNEY(+)	-	1	↓
272.0946_5.57	[M+H]+	Deoxycytidine	227.090607	C9H13N3O4	0.006887	1	0.858	KIDNEY(+)	[+CO2]	1	↓
273.0901_1.5	[M+Na]+	5-Methoxytryptophan	265.112308	C12H17N4OS	0.000718	1	0.758	KIDNEY(+)	[+O]	0.95	↓
278.0404_3.57	[M+H]+	Dopamine 4-sulfate	233.035796	C8H11NO5S	0.007498	1	0.641	KIDNEY(+)	[+CO2]	0.95	↑ dose-dependent
307.0513_2.24	[M+Na]+	D-Glucurono-6,3-lactone	176.03209	C6H8O6	-0.002377	1	0.698	KIDNEY(+)	[+C5H4N2O]	0.95	↑
348.0774_1.6	[M+H]+	2'-Deoxyguanosine 5'-monophosphate	347.063088	C10H14N5O7P	0.007036	1	0.713	KIDNEY(+)	-	1	↓

371.1139_1.84	[M+K]+	4-Hydroxynonenal	156.11503	C9H16O2	0.003618	0.977	0.666	KIDNEY(+)	[+C6H8O6]	1	↓
520.3444_8.18	[M+H]+	LysoPC(18:2(9Z,12Z))	519.332492	C26H50NO7P	0.004632	1	0.519	KIDNEY(+)	-	1	↑
113.0361_1.09	[M+H]+	Dihydrouracil	126.042928	C5H6N2O2	0.001546	1	0.122	LIVER(+)	[-CH2]	0.92	↓
145.0505_1.91	[M+H]+	3-Methylglutaconic acid	144.04226	C6H8O4	0.001064	1	0.777	LIVER(+)		0.92	↑
149.1182_5.93	[M+H]+	3-Hydroxyisoheptanoic acid	146.094295	C7H14O3	0.000979	1	0.221	LIVER(+)	[+H2]	0.92	↓
188.0695_2.69	[M+H]+	Indoleacrylic acid	187.063329	C11H9NO2	-0.001205	1	0.725	LIVER(+)	-	0.92	↑
241.1771_15.35	[M+K]+	Spermine	202.215746	C10H26N4	-0.001808			LIVER(+)	-	0.72	↓ dose-dependent
254.0957_12.56	[M+Na]+	Suberylglycine	231.110674	C10H17NO5	-0.004198	1	0.689	LIVER(+)	-	0.72	↓ dose-dependent
256.0920_3.83	[M+K]+	Propionylcarnitine		C10H19NO4	-0.003	1	0.59	PLASMA(+)	-		↑
278.0352_3.57	[M+H]+	L-DOPA sulphate		C9H11NO7S	0.002	1	0.65	PLASMA(+)	-		↑
254.0949_5.04	[M+H]+	Neopterin		C9H11N5O4	0.006	1	0.75	PLASMA(+)	-		↑
229.1555_1.65	[M+H]+	L-isoleucyl-L-proline		C11H20N2O3	0.0001	1	0.79	PLASMA(+)	-		↑
222.0898_6.78	[M-H]-	5-Methyldeoxycytidine		C10H15N3O4	0001	1	0.69	PLASMA(-)	[-H2O]		↑
239.1498_4.34	[M+H]+	Homoanserine		C11H18N4O3	-0.0005	1	0.7	PLASMA(+)	[-O]		↓
515.1538_6.72	[M-H]-	S-Adenosylhomocysteine		C14H20N6O5S		1	0.74	PLASMA(-)	[+C5H8O4]		↑
264.0569_3.23	[M+H]	N-acetyl-S-(3-oxo-3-carboxynpropyl) cysteine	263.0464	C9H13NO6S	0.0059	1	1	LIVER(+)	[-H2]	1	↑
154.0705	[M+H]+	Dopamine	153.0790	C8H11NO2	0.0016	1	0.77	LIVER(+)	-	0.79	↑
206.1405_10.45	[M+NH4]+	Nonic acid	188.1049	C9H16O4	0.0008	1	0.76	LIVER(+)	-	1	↓

7.5. Discussion

This study was designed to enrich the existing knowledge of the metabolic alterations related to the CMS. So far, most of the *in vivo* studies have used significantly high CMS doses to invoke the drug's toxicity. For instance, the most recent experiment of Nguyen et. al., induced *in vivo* CMS toxicity by administering 25 and 50 mg/kg [128] as low and high dose respectively, which is 10 folds higher than the recommended human dose (2.5-5 mg/kg/day). The high dosing approaches provide ample information on biochemical alterations occurring in extremely toxic conditions offering clear evidence of the triggered pathways leading to the observed clinical symptoms. On the other hand, such dosing regimens are never met in the clinical setting, where the administered drug levels are strictly regulated and immediately taken care of, either by lowering or even by stopping the administration. However, CMS remains the last resort antibiotic for patients infected by multi-drug-resistant bacteria and thus, the metabolic dysregulations caused even by the normal dosing schemes should be investigated. Anticipating the early and frequently latent biochemical alterations provoked by the drug, will provide the ability to adjust the curation protocol by reducing/ stopping the administration. Under this notion, the current study attempted to simulate *in vivo*, the metabolic alterations exhibiting by lower doses of CMS, that are comparable to those administered to humans. Thus, 15 mice were separated in three equal groups and received 0 mg/kg (C - the control group), 1 mg/kg (LD) and 1.5 mg/kg (HD) of CMS. The study involved the plasma, the kidneys, and the livers, to investigate the latent background of colistin nephrotoxicity, and to examine the impact of the drug in the circulatory system and in the liver as well, as the knowledge on the latter is limited.

An RP-HRMS-based metabolomics protocol was employed to analyse the samples, whereas an array of univariate and multivariate methodologies were combined for the statistical process. The PCA and OPLS-DA focused on the pairwise comparisons of C-LD and LD-HD groups. PCA showed that the most pronounced metabolic perturbation occurred by the LD to the renal tissue. This implies that the drug triggers important metabolic alterations in kidneys, even at the lower doses, when there are no clinical data that could indicate toxicity.

The OPLS-DA models classified efficiently the C-LD and LD-HD groups, providing satisfactory figures of merit.

Interestingly, the liver dataset offered the highest number of important features in both OPLS-DA and ROC curve analysis and the most of them were downregulated with the increase of the drug. However, only 9 of the liver's differentiated variables were finally identified, speculating that the drug impairs the liver metabolism, dysregulating metabolites derivatives that remain unknown. In addition, the PLS models proved the existence of linear correlations between the dose level and the alterations expressed in mice metabolomic profiles. The PLS models were only used to verify the linear correlation of metabolomics profile and CMS dose. The PLS results were not exploited for the variable selection, as PLS is a more complex model encompasses 3 groups, whereas the pairwise OPLS-DA model afford more interpretable results.

Concomitantly, the SUS-plot was used to investigate the existence of metabolites that are linearly correlated to the dose and highlighted 16. Six of the dose-correlated variables were identified: suberyglycine (liver, ↓, R^2= 0.75), spermine (liver, ↓, R^2= 0.87), (R)-N-methylsalsolinol (kidney, ↑, R^2= 0.79), phenylacetic acid (kidney, ↓, R^2= 0.85), 2,8-dihydroxyadenine (kidney, ↑, R^2= 0.82), dopamine 4-sulfate (kidney, ↑, R^2= 0.78), and examples of their box-plots are shown in Figure 35.

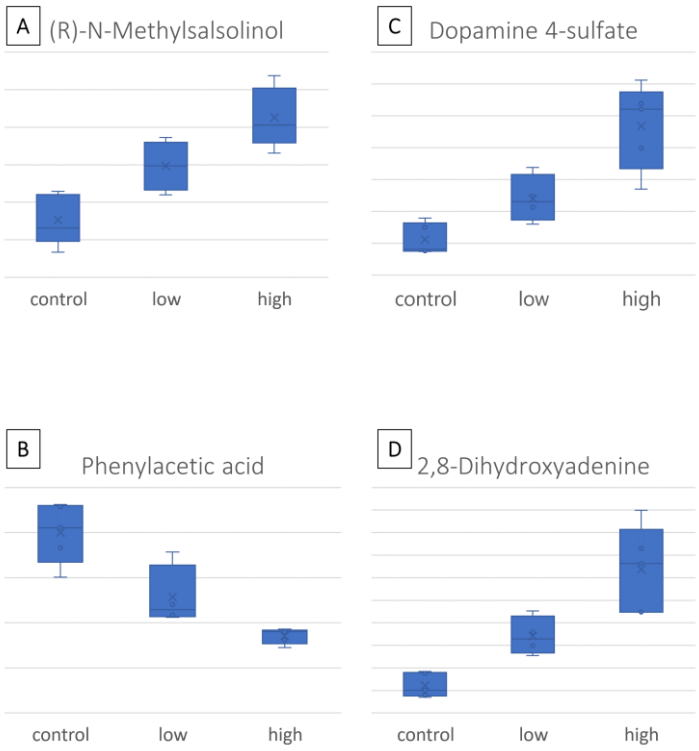

Figure 36 Box-plots of the most important CMS dose-responding metabolites: (A) N-Methylsalsolinol, detected in kidney; (B) Dopamine-4-sulfate, detected in kidney; (C) Phenylacetic acid, detected in kidney; (D) 2,8- Dihydroxyadenine, detected in kidney.

7.5.1. Alterations of dopamine pathway

The administration of CMS provoked changes in 4 metabolites that belong to the dopamine biochemical pathways, i.e., the dopamine (DA, liver, ↑), the domapine-4-sulfate (DA-4-S, kidney, dose-increased), the L-DOPA sulfate (L-DOPA-S, plasma ↑) and the N-methyl-R-salsolinol (M*N*RSal, kidney, dose-increased). The M*N*RSal that showed elevated response to the drug dose, is an endogenous neurotoxin, related to cell apoptosis. M*N*RSal is the enzymatic product of R-salsolinon (R-Sal) which in turn is formed by DA under the action of R-Sal synthases. M*N*RSal has been detected in urine of patients with Parkinson disease and is considered more toxic than R-Sal [130]. M*N*RSal presents apoptotic action, and it is suggested that the toxin impairs the mitochondrial permeability transition (PT) by reducing the mitochondrial membrane potential, resulting in increased release of apoptotic factors, as cytochrome *c*, into the cytoplasm. Also, the toxin activates the caspase-3 which also induces cell death [131]. Furthermore, M*N*RSal degradation products inhibit the mitochondrial complex-I causing apoptosis and increase the ROS as well[130]. The increased trend of the toxin in kidney, following the increase of

CMS, indicated early apoptosis in renal cells that could lead to severe kidney injury.

DA-4-S showed increased levels, in response to the drug dose in the kidney whereas L-DOPA-S was increased in plasma as well. The sulfonation locus of the endogenous/ exogenous phenols and catechols, i.e., DA and L-DOPA, probably happens in the upper gastrointestinal track, where the responsible enzymes are mainly expressed [132]. It should be noted that the sulfated forms of DA are predominant in human blood and represent the 90% of the total DA [132]. Furthermore, the sulfonation of DA is pivotal for metabolites binding with its receptors [132]. The increased levels of DA-4-S and L-DOPA-S in plasma and kidneys suggest elevated biosynthesis of DA. DA is a natriuretic hormone and regulates the sodium levels, inducing sodium excretion and constraining its reabsorption at the proximal tubule [133]. Thus excessive action of DA leads to limited levels of circulatory sodium, condition which is linked with the occurrence of hypotension [134]. Besides, it has been speculated that impairment of the estimated glomerular filtration rate (eGRF) is associated with neurological adverse effects, i.e., the limited eGRF leads to increased circulation levels of uremic toxins and kidney hormones that end up in the dopaminergic system of the brain [135]. Furthermore, DA is associated with the liver fatty acid (FA) metabolism occurring in the mitochondria of hepatocytes. The regulation of FA metabolism depends on the expression of carnitine palmitoyl transferase (CPT) I and CPT II. The catecholamines as DA induce the CPT gene expression in the hepatocytes, inducing ketogenesis. Thus, the increase of DA in CMS mice indicates dysregulation in liver FA metabolism[136]. The increased ketogenesis is linked to hypoglycemia. The acute increase of ketones leads to nausea, vomiting, pain, lethargy and even come, whereas chronic ketosis can cause hepatic transaminase elevation [137].

7.5.2. Down-regulation of renal phenylacetic acid

Phenylacetic acid (PAA) exhibited the opposite response to the drug dose, in the renal tissue i.e., metabolite decreasing levels in response to CMS increase. PAA is produced by phenylalanine degradation and is considered as a uremic toxin. The circulatory levels of PAA have been increased in patients with chronic

kidney disease (CKD) [138]. Besides, the metabolite inhibits the expression of the inducible nitric oxide synthase (iNOS) in mononuclear leukocytes in end-stage renal failure patient [139]. iNOS is expressed when the cells are triggered by proinflammatory cytokines and produce nitric oxide (NO) as a critical response of the immune system. In the current case, the PAA was found decreased by the CSM, speculating failure to balance the expression of iNOS leading to elevated levels of NO in kidney. The overexpression of iNOS, is linked to a variety of human diseases as the septic shock and pain [140]. Furthermore, the decreased renal PAA-levels suggest kidney impairment resulting in the limitation of renal filtration ability. Thus, the circulatory substances do not pass from blood to kidneys but accumulate in the circulation.

7.5.3. Up-regulation of renal 2,8-dihydroxyadenine

The accumulation of 2,8-dihydroxyadenine (2,8-DHA) renal levels by CMS dosing, is a worth mentioning finding. 2,8-DHA is an adenine metabolite, accumulated in cases of adenine phosphoribosyl-transferase (APRT) deficiency, which is a rare autosomal metabolic disorder, associated to uric acid's metabolism. 2,8-DHA exhibits low solubility, thus, its overexpression leads to formation and precipitation of urinary crystals and kidney stones, leading to urolithiasis or nephropathy [141].

7.5.4. Down-regulation of liver suberyglycine

Suberylglycine decreased by CMS administration in the liver tissue. There is limited literature concerning this substance. The metabolite is normally occurring as a product of fatty acid metabolism and is formed through the oxidation of suberyl-CoA, an intermediate of the fatty acid metabolism. Suberylglycine is primarily associated with a group of inherited metabolic disorders known as organic acidemias, leading to the accumulation of various organic acids, including suberylglycine, in the body. The measurement of suberylglycine levels in biological samples, such as urine or blood, are used as a diagnostic marker for certain organic acidemias [142],[143]. The elevated urinary levels of suberyglycine have been associated with hereditary medium-chain acyl-CoA hydrogenase (MCAD) failing [142], however, in the current case, metabolite's levels were detected decreased in the liver. MCAD is suggested to be the most usual cause of nonalcoholic fatty liver disease

(NAFLD . Paula et. al., have reported a rare hepatic steatosis secondary to chronic case, expressed in an infant, where the levels of circulatory suberyglycine were elevate[144]. This early observation of potential CSM-induced MCAD deficiency is key evidence and should be further investigated.

7.5.5. Down-regulation of liver spermine

Spermine was also found to be downregulated in liver, by the increase of CMS. The compound is a polyamine, naturally present in cells and tissues of living organisms, including humans. It is derived from the amino acid ornithine through a series of enzymatic reactions. Spermine plays important roles in various biological processes, including cell growth, proliferation, and DNA stabilization [145]. It has been found that, in plasma of patients with chronic renal failure, the circulatory levels of spermine are decreased [146]. In our study, the hepatic levels of spermine were decreased with the increased CMS dose. The polyamines as spermine are proved to counterbalance drug adverse effects, as the hepatotoxicity, and thus are administrated as protective agents [145], inhibiting cell apoptosis [147].

7.5.6. Altered purine metabolism and renal dysfunction.

As mentioned above, all the identified metabolites were submitted to a naïve pathway analysis, showing extended dysregulation in the renal purine metabolism. Therefore, the metabolites participating in purine metabolism were targeted determined in the raw MS data of kidney samples (C, LD, and HD). The results were used for additional pathway analysis, based on semi-quantitative data. This step showed extensive alterations to the purine metabolism (17/65 altered metabolites, P value (FDR) = 0.04) between the C and LD groups, whereas the increase of the dose (LD-HD) did not affect the observed dysregulation. The results of the semi-quantitative pathway analysis of C-LD groups are described in Figure 36.

The 17 altered metabolites were found to be increased by the administration of CMS, with the main perturbations occurring in the sub-pathways that result in the formation of xanthine-uric acid (UA) and guanine, Figure 6. The dysregulation of nine sequential metabolites, in the sub-pathway of UA formation, provides strong proof that CMS causes severe effects on renal purines metabolism.

Purines are enzymatically transformed into hypoxanthine and then into xanthine, which is the precursor of UA. The formation of UA is followed by the generation of superoxide anions and reactive oxygen species (ROS), as degradation byproducts. The extended production of UA, as observed in the current case, leads to high levels of intracellular oxidative-stress-inducing-factors causing cell damage [148]. The dysregulation of purine metabolism, and particularly the hyperuricemia (elevation of UA) is associated with kidney injury and is a marker for the progression of CKD [149]. There are several proposed associations between UA and kidney impairment: i) formation of monosodium urate crystals that precipitate in the tubules of the extra-renal system; ii) oxidative stress due to the intracellular pro-oxidative properties of UA that cause endothelial dysfunction, renal fibrosis, inflammation and glomerulosclerosis; iii) UA prevents the nitric oxide (NO) synthesis and thus hinder the endothelial cell proliferation [148],[149]. The increased levels of xanthine and UA are also associated with the aging-induced renal impairment [149].

Besides, the alterations of purine metabolism could be the key-point regarding the connection of neuro- and nephro-toxicity occurring due to CMS administration. It is proposed that the decreased eGRF leads to increase of circulatory-levels of metabolic waste (uremic toxins and kidney hormones) provoking all types neurological complications by i.e., triggering the nervus system (dopaminergic system) causing brain dysfunction [135],[150]. In addition, there are several reports of cognitive impairment (CI) in patients with CKD [150]. A recent computational docking study tested the binding affinity of xanthine, hypoxanthine and 2,8-DHA with acetylcholinesterase (AChE) and, the UA showed the higher binding affinity whereas xanthine and hypoxanthine presented high docking scores as well [150]. AChE hydrolyses acetylcholine that is important for the learning and the memory, and so, its rapid degradation by the AChE leads to dementia and CI [151]. *In-vitro* testing showed that hypoxanthine enhanced the action of AChE [151].

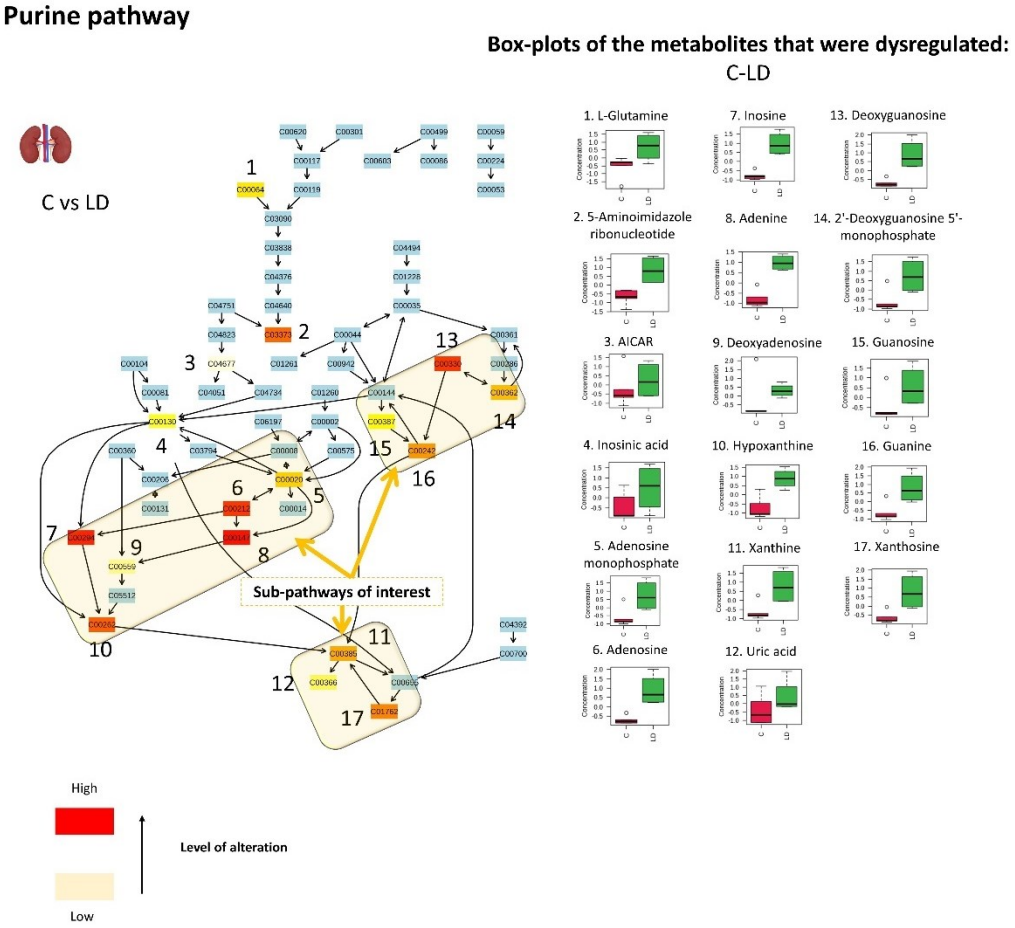

Figure 37 Graphical description of the alteration occurred in renal purine metabolism with the administration of 1.0 mg/kg CMS. The sub-pathways of interest point out the locations that sequential alteration were observed. The boxplots show the content of the perturbated metabolites in C (red) and LD (green) groups.

7.6. Materials and methods

The study employed plasma, kidney, and liver samples of 15 C57Bl/6 (weight: 20-25 g; age: 8–10 weeks) mice. All in vivo experiments were carried out in accordance with the “Guide for the care and use of Laboratory animals” and experiments were approved by the Ethics Committee (Approval No: 574234/20-07-2020). The mice were housed and maintained according to the ARRIVE guidelines. The number of animals needed to achieve statistical power > 80 % has been calculated using GPower 3.1. The animals were randomized in three groups (n = 5 for each group) as follows: (i) Control (NaCl 0.9%), (ii) LD (CSM 1 mg/kg/day), (iii) HD (CSM 1.5 mg/kg/day). The drug and the normal saline respectively, were injected intra muscularly (i.m.) to the thigh of each laboratory animal. The administration has been repeated for 5 sequential days and the sixth day the laboratory animals were sacrificed.

For the plasma extraction, 200 μL of the sample were mixed with 600 μL of frozen MeOH, centrifuged using a NEYA 16R centrifugation apparatus (REMI, Mumbai, India) at 10,000× g rpm, 5 min, 4 °C and the supernatant was stored at -80 oC. For the extraction of liver and kidney samples, 100 mg of the tissue were homogenized with 1 mL of MeOH-H2O (1:1, v/v) solution. A CRYOLYS EVOLUTION tissue homogenizer (Bertin Instruments, Rockville, MD, USA) and the homogenizing CKMix lysing kit (Bertin Corp., Rockville, MD, USA) were employed for a 2-step procedure. At first 500 μL of the solution were added in the falcon with the tissue and the "hard" mode (9600× g rpm, three 20 s cycles followed by 60 s pause) of the homogenizer was applied. Then the blend was centrifuged, and the supernatant was obtained. The rest of the solution was added in the homogenizing tube with the tissue remained and submitted to a second cycle of a "soft" mode (5000× g rpm, one 60 s cycle) homogenization. 400 μL and 500 μL of plasma and tissues extract respectively were evaporated to dryness by a HyperVAC-LITE centrifugal vacuum concentrator (Hanil Scientific Inc., Gimpo, Korea), and the remaining solid was reconstituted with 150 μL of IS mix solution, containing 1 ppm of yohimbine and reserpine in MeOH-H2O (1:1, v/v). The samples were analyzed using a Dionex UltiMate 3000 RSLC (Thermo Fischer Scientific, Dreieich, Germany) UPLC system, coupled to a Maxis Impact QTOF mass spectrometer (Bruker Daltonics, Bremen, Germany) that was equipped with an Electrospray Ionization source. The chromatographic column used was Acclaim RSL C18 column (2.1 x 100 mm, 2.2 μm, Thermo Fischer Scientific) and the elution of the analytes was performed with gradient conditions, by the ramp increase of the organic mobile phase. The DIA methodology (bbCID mode, in Bruker terminology) was selected of MS acquisition. A detailed description of the UPLC conditions and the parameters of ESI and MS are included in the Supplementary Materials, §4.1 & 4.2.

7.7. Conclusion

CMS is a last resort antibiotic factor administered for the treatment of infections caused by multidrug-resistant bacteria. However, the use of the drug is attenuated by the occurrence of neurological and renal complications resulting by its administration. Several studies have focused on CMS-induced neuro- and nephrotoxicity by administrating high doses of the drug to simulate the toxic condition. The current study aimed to shed light in the biochemical alterations triggered by the recommended human-doses, 1-1.5 mg/kg/day. Besides, the study investigated the impact of the drug on the circulatory system, and in the renal and liver function as well. So far, despite the existence of indications regarding the CMS-induce hepatotoxicity, there are no evidence for the displaying mechanisms. The study showed that even the lower human-dose (1 mg/kg) had severe impact on kidneys and also pointed out linear response between the drug-dose and the metabolic alterations for plasma, kidney, and liver. Sixteen variables showed significant correlation to the dose and 6 of them were identified: suberyglycine (liver, ↓), spermine (liver, ↓), (R)-N-methylsalsolinol (kidney, ↑), phenylacetic acid (kidney, ↓), 2,8-dihydroxyadenine (kidney, ↑), dopamine 4-sulfate (kidney, ↑). Summarizing, the results of the current study showed that CMS:

- Induces the renal dopamine pathway.
- Increases the renal levels of 2,8-DHA, and probably leads to the formation and precipitation of urinary crystals and kidney stones.
- Perturbates the renal purine metabolism, increasing the formation of xanthine, hypoxanthine, and UA. Xanthine is considered as AChE activator, leading to rapid degradation of achetylcholine. This is strong evidence for the share metabolic background of CMS induced nephrotoxicity and neurotoxicity.
- Perturbation of hepatic MCAD, probably leading to hepatic steatosis.
- Decrease of hepatic levels of spermine, which counterbalances hepatotoxicity by inhibiting the cell death.

References

[1] M. Schlander, K. Hernandez-Villafuerte, C. Y. Cheng, J. Mestre-Ferrandiz, and M. Baumann, “How Much Does It Cost to Research and Develop a New Drug? A Systematic Review and Assessment,” *Pharmacoeconomics*, vol. 39, no. 11, pp. 1243–1269, 2021.

[2] J. M. Banda, L. Evans, R. S. Vanguri, N. P. Tatonetti, P. B. Ryan, and N. H. Shah, “Data descriptor: A curated and standardized adverse drug event resource to accelerate drug safety research,” *Sci. Data*, vol. 3, pp. 1–11, 2016.

[3] G. A. Van Norman, “Limitations of Animal Studies for Predicting Toxicity in Clinical Trials: Is it Time to Rethink Our Current Approach?,” *JACC Basic to Transl. Sci.*, vol. 4, no. 7, pp. 845–854, 2019.

[4] C. H. C. Leenaars *et al.*, “Animal to human translation: A systematic scoping review of reported concordance rates,” *J. Transl. Med.*, vol. 17, no. 1, pp. 1–22, 2019.

[5] R. S. Thomas *et al.*, “The US Federal Tox21 Program: A strategic and operational plan for continued leadership,” *ALTEX*, vol. 35, no. 2, pp. 163–168, 2018.

[6] A. B. Raies and V. B. Bajic, “In silico toxicology: computational methods for the prediction of chemical toxicity,” *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 6, no. April, pp. 147–172, 2016.

[7] J. Hemmerich and G. F. Ecker, “In silico toxicology: From structure–activity relationships towards deep learning and adverse outcome pathways,” *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 10, no. 4, pp. 1–23, 2020.

[8] S. A. Langhans, “Three-Dimensional in Vitro Cell Culture Models in Drug Discovery and Drug Repositioning,” vol. 9, no. January, pp. 1–14, 2018.

[9] M. Sun *et al.*, “3D Cell Culture — Can It Be As Popular as 2D Cell Culture ?,” 2021.

[10] A. M. Handorf *et al.*, “Tissue Stiffness Dictates Development , Homeostasis , and Disease Progression Tissue Stiffness Dictates Development , Homeostasis , and Disease Progression,” vol. 6278, 2015.

[11] A. Liu, S. Seal, H. Yang, and A. Bender, “Using chemical and biological data to predict drug toxicity,” *SLAS Discov.*, vol. 28, no. 3, pp. 53–64, 2023.

[12] P. Joseph, “Transcriptomics in toxicology.,” *Food Chem. Toxicol. an Int. J. Publ. Br. Ind. Biol. Res. Assoc.*, vol. 109, no. Pt 1, pp. 650–662, Nov. 2017.

[13] A. M. Araújo, F. Carvalho, P. G. De Pinho, and M. Carvalho, “Toxicometabolomics: Small molecules to answer big toxicological questions,” *Metabolites*, vol. 11, no. 10, 2021.

[14] M. R. Boland, A. Jacunski, T. Lorberbaum, J. D. Romano, R. Moskovitch, and N. P. Tatonetti, “Systems biology approaches for identifying adverse drug reactions and elucidating their underlying biological mechanisms.,” *Wiley Interdiscip. Rev. Syst. Biol. Med.*, vol. 8, no. 2, pp. 104–122, 2016.

[15] A. E. Steuer, L. Brockbals, and T. Kraemer, “Metabolomic strategies in biomarker research-new approach for indirect identification of drug consumption and sample manipulation in clinical and forensic toxicology?,” *Front. Chem.*, vol. 7, no. MAY, 2019.

[16] R. D. Beger, M. A. Schmidt, and R. Kaddurah-Daouk, “Current concepts in pharmacometabolomics, biomarker discovery, and precision medicine,” *Metabolites*, vol. 10, no. 4, 2020.

[17] J. Nicholson, H. Keun, and T. Ebbels, “COMET and the challenge of drug safety screening,” *J. Proteome Res.*, vol. 6, no. 11, pp. 4098–4099, 2007.

[18] F. R. Pinu *et al.*, “Systems biology and multi-omics integration: Viewpoints from the metabolomics research community,” *Metabolites*, vol. 9, no. 4, pp. 1–31, 2019.

[19] J. Holle *et al.*, “Inflammation in Children with CKD Linked to Gut Dysbiosis and Metabolite Imbalance,” *J. Am. Soc. Nephrol.*, vol. 33, no. 12, pp. 2259–2275, 2022.

[20] V. Tolstikov, A. James Moser, R. Sarangarajan, N. R. Narain, and M. A. Kiebish, “Current status of metabolomic biomarker discovery: Impact of study design and demographic characteristics,” *Metabolites*, vol. 10, no. 6, pp. 1–12, 2020.

[21] A. V. Aderemi, A. O. Ayeleso, O. O. Oyedapo, and E. Mukwevho, “Metabolomics: A scoping review of its role as a tool for disease biomarker discovery in selected non-communicable diseases,” *Metabolites*, vol. 11, no. 7, 2021.

[22] W. B. Dunn *et al.*, “Molecular phenotyping of a UK population : defining the human serum metabolome,” pp. 9–26, 2015.

[23] S. Moco, “Studying Metabolism by NMR-Based Metabolomics,” *Front. Mol. Biosci.*, vol. 9, no. April, pp. 1–12, 2022.

[24] A. H. Emwas *et al.*, “Nmr spectroscopy for metabolomics research,” *Metabolites*, vol. 9, no. 7, 2019.

[25] D. Tang, L. Zou, X. Yin, and C. N. Ong, “HILIC-MS FOR METABOLOMICS : AN ATTRACTIVE AND COMPLEMENTARY APPROACH TO RPLC-MS,” pp. 1–27, 2014.

[26] C. S. Ho *et al.*, “Electrospray Ionisation Mass Spectrometry : Principles and Clinical Applications,” vol. 24, no. February, pp. 3–12, 2003.

[27] Ulrich Boesl, “TIME-OF-FLIGHT MASS SPECTROMETRY: INTRODUCTION TO THE BASICS,” 2016.

[28] D. R. Allen and B. C. McWhinney, “Quadrupole Time-of-Flight Mass Spectrometry: A Paradigm Shift in Toxicology Screening Applications,” *Clin. Biochem. Rev.*, vol. 40, no. 3, pp. 135–146, 2019.

[29] R. Wang, Y. Yin, and Z. J. Zhu, “Advancing untargeted metabolomics using data-independent acquisition mass spectrometry technology,” *Anal. Bioanal. Chem.*, vol. 411, no. 19, pp. 4349–4357, 2019.

[30] D. S. Wishart *et al.*, “HMDB 4.0: The human metabolome database for 2018,” *Nucleic Acids Res.*, vol. 46, no. D1, pp. D608–D617, 2018.

[31] C. A. Smith *et al.*, “METLIN: a metabolite mass spectral database.,” *Ther. Drug Monit.*, vol. 27, no. 6, pp. 747–751, Dec. 2005.

[32] H. Horai *et al.*, “MassBank: a public repository for sharing mass spectral data for life sciences.,” *J. Mass Spectrom.*, vol. 45, no. 7, pp. 703–714, Jul. 2010.

[33] C. Kuhl, R. Tautenhahn, C. Böttcher, T. R. Larson, and S. Neumann, “CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets.,” *Anal. Chem.*, vol. 84, no. 1, pp. 283–289, Jan. 2012.

[34] Y. M. Tikunov, S. Laptenok, R. D. Hall, A. Bovy, and R. C. H. de Vos, “MSClust: a tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data.,” *Metabolomics*, vol. 8, no. 4, pp. 714–718, Aug. 2012.

[35] C. D. Broeckling, F. A. Afsar, S. Neumann, A. Ben-Hur, and J. E. Prenni, “RAMClust: A novel feature clustering method enables spectral-matching-based annotation for metabolomics data,” *Anal. Chem.*, vol. 86, no. 14, pp. 6812–6817, 2014.

[36] R. R. Da Silva, P. C. Dorrestein, and R. A. Quinn, “Illuminating the dark matter in metabolomics,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 41, pp. 12549–12550, 2015.

[37] T. Huan, C. Tang, R. Li, Y. Shi, G. Lin, and L. Li, “MyCompoundID MS/MS Search: Metabolite Identification Using a Library of Predicted Fragment-Ion-Spectra of 383,830 Possible Human Metabolites,” *Anal. Chem.*, vol. 87, no. 20, pp. 10619–10626, 2015.

[38] K. Strimbu and J. A. Tavel, “What are biomarkers?,” *Curr. Opin. HIV AIDS*, vol. 5, no. 6, pp. 463–466, 2010.

[39] R. M. Califf, “Biomarker definitions and their applications,” *Exp. Biol. Med.*, vol. 243, no. 3, pp. 213–221, 2018.

[40] J. E. McDermott *et al.*, “Challenges in Biomarker Discovery: Combining Expert Insights with Statistical Analysis of Complex Omics Data.,” *Expert Opin. Med. Diagn.*, vol. 7, no. 1, pp. 37–51, Jan. 2013.

[41] A. Koulman, G. A. Lane, S. J. Harrison, and D. A. Volmer, “From differentiating metabolites to biomarkers,” *Anal. Bioanal. Chem.*, vol. 394, no. 3, pp. 663–670, 2009.

[42] N. R. Anwardeen, I. Diboun, Y. Mokrab, A. A. Althani, and M. A. Elrayess, “Statistical methods and resources for biomarker discovery using metabolomics,” *BMC Bioinformatics*, vol. 24, no. 1, pp. 1–18, 2023.

[43] K. Hajian-Tilaki, “Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation.,” *Casp. J. Intern. Med.*, vol. 4, no. 2, pp. 627–635, 2013.

[44] J. Wang *et al.*, “Identification of potential plasma biomarkers and metabolic dysfunction for unstable angina pectoris and its complication based on global metabolomics,” *Biosci. Rep.*, vol. 39, no. 3, pp. 1–10, 2019.

[45] L. Shi, J. A. Westerhuis, J. Rosén, R. Landberg, and C. Brunius, “Variable selection and validation in multivariate modelling,” *Bioinformatics*, vol. 35, no. 6, pp. 972–980, 2019.

[46] M. B. Kursa, “Feature Selection with the Boruta Package,” vol. 36, no. 11, 2010.

[47] S. Cacciatore, C. Luchinat, and L. Tenori, “Knowledge discovery by accuracy maximization,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 14, pp. 5117–5122, 2014.

[48] M. M. Zinga *et al.*, “KODAMA exploratory analysis in metabolic phenotyping,” *Front. Mol. Biosci.*, vol. 9, no. January, pp. 1–7, 2023.

[49] M. B. Kursa, “Boruta for those in a hurry,” pp. 1–6, 2020.

[50] Y. H. Yun *et al.*, “Informative metabolites identification by variable importance analysis based on random variable combination,” *Metabolomics*, vol. 11, no. 6, pp. 1539–1551, 2015.

[51] P. Efentakis *et al.*, “Mineralocorticoid Receptor Pathway Is a Key Mediator of Carfilzomib-induced Nephrotoxicity : Preventive Role of Eplerenone,” *Hemasphere*, 2022.

[52] I. Barla *et al.*, “An Untargeted Metabolomics Approach on Carfilzomib-Induced Nephrotoxicity,” *Molecules*, vol. 27, no. 22, 2022.

[53] P. Efentakis *et al.*, “Molecular mechanisms of carfilzomib-induced cardiotoxicity in mice and the emerging cardioprotective role of metformin,” 2019.

[54] E. P. Rhee, “A Systems-Level View of Renal Metabolomics,” vol. 38, no. 2, pp. 142–150, 2019.

[55] Z. Gao and X. Chen, “Fatty Acid β-Oxidation in Kidney Diseases: Perspectives on Pathophysiological Mechanisms and Therapeutic Opportunities,” *Front. Pharmacol.*, vol. 13, no. April, pp. 1–10, 2022.

[56] H. S. Jang, M. R. Noh, J. Kim, and B. J. Padanilam, “Defective Mitochondrial Fatty Acid Oxidation and Lipotoxicity in Kidney Diseases,” *Front. Med.*, vol. 7, no. March, pp. 1–8, 2020.

[57] S. K. Natarajan and J. A. Ibdah, “Role of 3-hydroxy fatty acid-induced hepatic lipotoxicity in acute fatty liver of pregnancy,” *Int. J. Mol. Sci.*, vol. 19, no. 1, pp. 1–17, 2018.

[58] S. Jeon, “Regulation and function of AMPK in physiology and diseases,” vol. 48, no. 9, pp. e245-13, 2016.

[59] M. H. Mahbub *et al.*, “Association of plasma branched-chain and aromatic amino acids with reduction in kidney function evaluated in apparently healthy adults,” *J. Clin. Med.*, vol. 10, no. 22, 2021.

[60] C. L. Palm, K. T. Nijholt, B. M. Bakker, and B. D. Westenbrink, “Short-Chain Fatty Acids in the Metabolism of Heart Failure – Rethinking the Fat Stigma,” *Front. Cardiovasc. Med.*, vol. 9, no. July, pp. 1–9, 2022.

[61] I. Armando, V. A. M. Villar, and P. A. Jose, “Dopamine and renal function and blood pressure regulation,” *Compr. Physiol.*, vol. 1, no. 3, pp. 1075–1117, 2011.

[62] A. M. Madella, J. Van Bergenhenegouwen, J. Garssen, R. Masereeuw, and S. A. Overbeek, “Microbial-Derived Tryptophan Catabolites, Kidney Disease and Gut Inflammation,” *Toxins (Basel).*, vol. 14, no. 9, pp. 1–21, 2022.

[63] A. Fiore and P. J. Murray, “Tryptophan and indole metabolism in immune regulation,” *Curr. Opin. Immunol.*, vol. 70, pp. 7–14, 2021.

[64] Y. Chen and G. J. Guillemin, “Kynurenine pathway metabolites in humans: Disease and healthy states,” *Int. J. Tryptophan Res.*, vol. 2, no. 1, pp. 1–19, 2009.

[65] I. Zakrocka and W. Załuska, “Kynurenine pathway in kidney diseases,” *Pharmacol. Reports*, vol. 74, no. 1, pp. 27–39, 2022.

[66] M. Ala and S. P. Eftekhar, “The Footprint of Kynurenine Pathway in Cardiovascular Diseases,” *Int. J. Tryptophan Res.*, vol. 15, 2022.

[67] M. Fujiwara *et al.*, “Formation of 5-hydroxykynurenine and 5-hydroxykynurenamine from 5-hydroxytryptophan in rabbit small intestine,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 76, no. 3, pp. 1145–1149, 1979.

[68] D. W. Pomfret, K. W. Schenck, P. Fludzinski, and M. L. Cohen, “Interaction of 5-hydroxykynurenamine, L-kynurenine and kynuramine with multiple serotonin receptors in smooth muscle.,” *J. Pharmacol. Exp. Ther.*, vol. 241, no. 2, pp. 465–471, May 1987.

[69] G. Kaur and P. Krishan, “Understanding Serotonin 5-HT2A Receptors-regulated cellular and molecular Mechanisms of Chronic Kidney Diseases,” *Ren. Replace. Ther.*, vol. 6, no. 1, pp. 1–11, 2020.

[70] C. C. Clement *et al.*, “3-hydroxy-L-kynurenamine is an immunomodulatory biogenic amine,” *Nat. Commun.*, vol. 12, no. 1, pp. 1–17, 2021.

[71] C. B. Wong, A. Tanaka, T. Kuhara, and J. Z. Xiao, “Potential effects of

indole-3-lactic acid, a metabolite of human bifidobacteria, on NGF-induced neurite outgrowth in PC12 cells," *Microorganisms*, vol. 8, no. 3, 2020.

[72] S. Krishnan *et al.*, "Gut Microbiota-Derived Tryptophan Metabolites Modulate Inflammatory Response in Hepatocytes and Macrophages.," *Cell Rep.*, vol. 23, no. 4, pp. 1099–1111, 2018.

[73] S. Wang *et al.*, "Identification of Biomarkers of Sepsis-Associated Acute Kidney Injury in Pediatric Patients Based on UPLC-QTOF/MS," *Inflammation*, vol. 43, no. 2, pp. 629–640, 2020.

[74] S. K. Manna *et al.*, "UPLC-MS-based urine metabolomics reveals indole-3-lactic acid and phenyllactic acid as conserved biomarkers for alcohol-induced liver disease in the Ppara-null mouse model," *J. Proteome Res.*, vol. 10, no. 9, pp. 4120–4133, 2011.

[75] H. Tsugawa *et al.*, "MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis," *Nat. Methods*, vol. 12, no. 6, pp. 523–526, 2015.

[76] H. Luan, F. Ji, Y. Chen, and Z. Cai, "statTarget: A streamlined tool for signal drift correction and interpretations of quantitative mass spectrometry-based omics data," *Anal. Chim. Acta*, vol. 1036, pp. 66–72, 2018.

[77] J. Xia, D. I. Broadhurst, M. Wilson, and D. S. Wishart, "Translational biomarker discovery in clinical metabolomics: An introductory tutorial," *Metabolomics*, vol. 9, no. 2, pp. 280–299, 2013.

[78] A. K. Smilde, J. J. Jansen, H. C. J. Hoefsloot, R. J. A. N. Lamers, J. van der Greef, and M. E. Timmerman, "ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data," *Bioinformatics*, vol. 21, no. 13, pp. 3043–3048, 2005.

[79] Y. C. Tai and T. P. Speed, "A multivariate empirical Bayes statistic for replicated microarray time course data," *Ann. Stat.*, vol. 34, no. 5, pp. 2387–2412, 2006.

[80] Z. Lai *et al.*, "Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics," *Nat. Methods*, vol. 15, no. 1, pp. 53–56, 2018.

[81] L. Li *et al.*, "MyCompoundID: Using an evidence-based metabolome library for metabolite identification," *Anal. Chem.*, vol. 85, no. 6, pp. 3401–3408, 2013.

[82] M. B. Morelli *et al.*, "Cardiotoxicity of Anticancer Drugs: Molecular Mechanisms and Strategies for Cardioprotection," *Front. Cardiovasc. Med.*, vol. 9, no. April, 2022.

[83] K. Ananthan and A. R. Lyon, "The Role of Biomarkers in Cardio-Oncology," *Journal of Cardiovascular Translational Research*, vol. 13, no. 3. Springer, pp. 431–450, 01-Jun-2020.

[84] L. Michel *et al.*, “Cardiac biomarkers for the detection of cardiotoxicity in childhood cancer—a meta-analysis,” *ESC Hear. Fail.*, vol. 7, no. 2, pp. 423–433, 2020.

[85] A. Linares Ballesteros *et al.*, “Early-onset Cardiotoxicity assessment related to anthracycline in children with leukemia. A Prospective Study,” *Colomb. medica (Cali, Colomb.*, vol. 52, no. 1, p. e2034542, 2021.

[86] X. Yang *et al.*, “Potential Gene Association Studies of Chemotherapy-Induced Cardiotoxicity: A Systematic Review and Meta-Analysis,” *Front. Cardiovasc. Med.*, vol. 8, no. June, pp. 1–15, 2021.

[87] S. H. Shah and C. B. Newgard, “Integrated Metabolomics and Genomics,” *Circ. Cardiovasc. Genet.*, vol. 8, no. 2, pp. 410–419, 2015.

[88] Schrimpe-R, A. C. Utledge, S. G. Codreanu, S. D. Sherrod, and J. A. Mclean, “Untargeted Metabolomics Strategies — Challenges and Emerging Directions,” 2016.

[89] W. G. Hunter *et al.*, “Metabolomic Profiling Identifies Novel Circulating Biomarkers of Mitochondrial Dysfunction Differentially Elevated in Heart Failure With Preserved Versus Reduced Ejection Fraction: Evidence for Shared Metabolic Impairments in Clinical Heart Failure,” *J. Am. Heart Assoc.*, vol. 5, no. 8, 2016.

[90] J. Ritterhoff and R. Tian, “Metabolismin cardiomyopathy: Every substrate matters,” *Cardiovasc. Res.*, vol. 113, no. 4, pp. 411–421, 2017.

[91] R. J. A. Wanders, J. Komen, and S. Kemp, “Fatty acid omega-oxidation as a rescue pathway for fatty acid oxidation disorders in humans,” *FEBS J.*, vol. 278, no. 2, pp. 182–194, 2011.

[92] I. M. S. Guerra *et al.*, “Mitochondrial Fatty Acid β-Oxidation Disorders: From Disease to Lipidomic Studies—A Critical Review,” *Int. J. Mol. Sci.*, vol. 23, no. 22, p. 13933, 2022.

[93] C. Beuchel *et al.*, “Whole Blood Metabolite Profiles Reflect Changes in Energy Metabolism in Heart Failure,” *Metabolites*, vol. 12, no. 3, pp. 1–16, 2022.

[94] Mamer and Reimer L, “On the mechanisms of the formation of L-alloisoleucine and the 2-hydroxy-3-methylvaleric acid stereoisomers from L-isoleucine in maple syrup urine disease patients and in normal humans,” no. 2, 1992.

[95] M. E. Brosnan and J. T. Brosnan, “Histidine Metabolism and Function,” 2020.

[96] A. J. Gianottil *et al.*, “Selective Killing of KZebsieZZa pneumoniae by,” vol. 265, no. 2, pp. 831–837, 1990.

[97] S. E. Reuter and A. M. Evans, “Carnitine and acylcarnitines: Pharmacokinetic, pharmacological and clinical aspects,” *Clin. Pharmacokinet.*, vol. 51, no. 9, pp. 553–572, 2012.

[98] T. Ahmad *et al.*, “Prognostic implications of long-chain acylcarnitines in heart failure and reversibility with mechanical circulatory support,” *J. Am. Coll. Cardiol.*, vol. 67, no. 3, pp. 291–299, 2016.

[99] Z. Bai and Z. Wang, “Genistein protects against doxorubicin-induced cardiotoxicity through Nrf-2/HO-1 signaling in mice model,” *Environ. Toxicol.*, vol. 34, no. 5, pp. 645–651, 2019.

[100] J. J. Guo *et al.*, “Alginate oligosaccharide prevents acute doxorubicin cardiotoxicity by suppressing oxidative stress and endoplasmic reticulum-mediated apoptosis,” *Mar. Drugs*, vol. 14, no. 12, pp. 1–13, 2016.

[101] S. Dalleau, M. Baradat, F. Guéraud, and L. Huc, “Cell death and diseases related to oxidative stress:4-hydroxynonenal (HNE) in the balance,” *Cell Death Differ.*, vol. 20, no. 12, pp. 1615–1630, 2013.

[102] J. P. Castro, T. Jung, T. Grune, and W. Siems, “4-Hydroxynonenal (HNE) modified proteins in metabolic diseases,” *Free Radic. Biol. Med.*, vol. 111, no. November 2016, pp. 309–315, 2017.

[103] M. Csala *et al.*, “On the role of 4-hydroxynonenal in health and disease,” *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1852, no. 5, pp. 826–838, 2015.

[104] R. S. Fletcher and G. G. Lavery, “The emergence of the nicotinamide riboside kinases in the regulation of NAD+ metabolism,” *J. Mol. Endocrinol.*, vol. 61, no. 3, pp. R107–R121, 2018.

[105] J. Clement, M. Wong, A. Poljak, P. Sachdev, and N. Braidy, “The Plasma NAD + Metabolome Is Dysregulated in ‘normal’ Aging,” *Rejuvenation Res.*, vol. 22, no. 2, pp. 121–130, 2019.

[106] M. Yuan, S. B. Breitkopf, X. Yang, and J. M. Asara, “A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue.,” *Nat. Protoc.*, vol. 7, no. 5, pp. 872–881, 2012.

[107] W. B. Dunn *et al.*, “Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry,” *Nat. Protoc.*, vol. 6, no. 7, pp. 1060–1083, 2011.

[108] X. Liu *et al.*, “LC-MS-Based Plasma Metabolomics and Lipidomics Analyses for Differential Diagnosis of Bladder Cancer and Renal Cell Carcinoma,” *Front. Oncol.*, vol. 10, no. May, pp. 1–14, 2020.

[109] T. Pluskal, S. Castillo, A. Villar-Briones, and M. Orešič, “MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data,” *BMC Bioinformatics*, vol. 11, 2010.

[110] A. M. De Livera *et al.*, “Statistical Methods for Handling Unwanted Variation in Metabolomics Data,” *Anal. Chem.*, vol. 87, no. 7, pp. 3606–3615, Apr. 2015.

[111] J. Fu *et al.*, “Optimization of metabolomic data processing using NOREVA,” *Nat. Protoc.*, vol. 17, no. 1, pp. 129–151, 2022.

[112] S. S. Mohapatra, S. K. Dwibedy, and I. Padhy, “Polymyxins, the last-resort antibiotics: Mode of action, resistance emergence, and potential solutions,” *J. Biosci.*, vol. 46, no. 3, Sep. 2021.

[113] M. J. Satlin *et al.*, “Clinical and Laboratory Standards Institute and European Committee on Antimicrobial Susceptibility Testing Position Statements on Polymyxin B and Colistin Clinical Breakpoints,” *Clin. Infect. Dis.*, vol. 71, pp. e523–e529, 2020.

[114] N. Grégoire, V. Aranzana-Climent, S. Magréault, S. Marchand, and W. Couet, “Clinical Pharmacokinetics and Pharmacodynamics of Colistin,” *Clin. Pharmacokinet. 2017 5612*, vol. 56, no. 12, pp. 1441–1460, May 2017.

[115] M. E. Falagas and S. K. Kasiakou, “Colistin: The Revival of Polymyxins for the Management of Multidrug-Resistant Gram-Negative Bacterial Infections,” *Clin. Infect. Dis.*, vol. 40, pp. 1333–41, 2005.

[116] P. J. Bergen, J. Li, C. R. Rayner, and R. L. Nation, “Colistin Methanesulfonate Is an Inactive Prodrug of Colistin against Pseudomonas aeruginosa,” *Antimicrob. Agents Chemother.*, vol. 50, no. 6, pp. 1953–1958, 2006.

[117] D. Plachouras *et al.*, “Population pharmacokinetic analysis of colistin methanesulfonate and colistin after intravenous administration in critically ill patients with infections caused by gram-negative bacteria,” *Antimicrob. Agents Chemother.*, vol. 53, no. 8, pp. 3430–3436, 2009.

[118] R. L. Nation *et al.*, “Framework for optimisation of the clinical use of colistin and polymyxin B: The Prato polymyxin consensus,” *Lancet Infect. Dis.*, vol. 15, no. 2, pp. 225–234, Feb. 2015.

[119] J. Rychlíčková, V. Kubíčková, P. Suk, and K. Urbánek, “Challenges of Colistin Use in ICU and Therapeutic Drug Monitoring: A Literature Review,” *Antibiot. 2023, Vol. 12, Page 437*, vol. 12, no. 3, p. 437, Feb. 2023.

[120] A. Haseeb *et al.*, “Dose Optimization of Colistin: A Systematic Review,” *Antibiotics*, vol. 10, no. 12, pp. 1–14, 2021.

[121] S. M. Garonzik *et al.*, “Population pharmacokinetics of colistin methanesulfonate and formed colistin in critically ill patients from a multicenter study provide dosing suggestions for various categories of patients,” *Antimicrob. Agents Chemother.*, vol. 55, no. 7, pp. 3284–3294, 2011.

[122] A. Inci, M. K. Toker, I. Guney Bicer, A. Derbent, and Z. Salihoglu, “Determination of colistin-related nephrotoxicity and risk factors in intensive care unit,” *North Clin Istanb*, vol. 5, no. 2, pp. 120–124, 2018.

[123] E. Ramezanzade, R. Ghanbari, and T. Yazdanipour, “Colistin-Induced Neurotoxicity in a Multidrug-Resistant UTI Patient with Cervical Cancer:

A Case Report," *Nephro-Urology Mon. 2023 151*, vol. 15, no. 1, p. 127122, Nov. 2023.

[124] C. Dai *et al.*, "Molecular Mechanisms of Neurotoxicity Induced by Polymyxins and Chemoprevention," *ACS Chem. Neurosci.*, vol. 10, no. 1, pp. 120–131, Jan. 2019.

[125] F. M. Alotaibi *et al.*, "Incidence and Risk Factors of Colistin-Induced Nephrotoxicity Associated with The International Consensus Guidelines for the Optimal Use of the Polymyxins: A Retrospective Study in a Tertiary Care Hospital, Saudi Arabia," *Antibiotics*, vol. 11, no. 11, 2022.

[126] K. Eljaaly *et al.*, "Colistin Nephrotoxicity: Meta-Analysis of Randomized Controlled Trials," *Open Forum Infect. Dis.*, vol. 8, no. 2, 2021.

[127] Z. Gai, S. L. Samodelov, G. A. Kullak-Ublick, and M. Visentin, "Molecular Mechanisms of Colistin-Induced Nephrotoxicity," *Molecules*, vol. 24, p. 653, 2019.

[128] N. P. Long *et al.*, "Delineation of the molecular mechanisms underlying Colistin-mediated toxicity using metabolomic and transcriptomic analyses," *Toxicol. Appl. Pharmacol.*, vol. 439, no. February, p. 115928, 2022.

[129] E. S. Jeong *et al.*, "Characterization of urinary metabolites as biomarkers of colistin-induced nephrotoxicity in rats by a liquid chromatography/mass spectrometry-based metabolomics approach," *Toxicol. Lett.*, vol. 248, pp. 52–60, Apr. 2016.

[130] Y. Cao *et al.*, "Neurotoxicity and underlying mechanisms of endogenous neurotoxins," *Int. J. Mol. Sci.*, vol. 22, no. 23, 2021.

[131] Y. Akao *et al.*, "Mitochondrial permeability transition mediates apoptosis induced by N-methyl(R)salsolinol, an endogenous neurotoxin, and is inhibited by Bcl-2 and rasagiline, N-propargyl-1(R)-aminoindan," *J. Neurochem.*, vol. 82, no. 4, pp. 913–923, 2002.

[132] K. Itäaho, S. Alakurtti, J. Yli-Kauhaluoma, J. Taskinen, M. W. H. Coughtrie, and R. Kostiainen, "Regioselective sulfonation of dopamine by SULT1A3 in vitro provides a molecular explanation for the preponderance of dopamine-3-O-sulfate in human blood circulation," *Biochem. Pharmacol.*, vol. 74, no. 3, pp. 504–510, 2007.

[133] M. R. Choi, "Renal dopaminergic system: Pathophysiological implications and clinical perspectives," *World J. Nephrol.*, vol. 4, no. 2, p. 196, 2015.

[134] S. Alshahrani, R. M. Rapoport, and M. Soleimani, "Vascular contractile reactivity in hypotension due to reduced renal reabsorption of Na+ and restricted dietary Na+," *Naunyn. Schmiedebergs. Arch. Pharmacol.*, vol. 390, no. 3, pp. 321–326, 2017.

[135] A. De Donato, V. Buonincontri, G. Borriello, G. Martinelli, and P. Mone, "The Dopamine System: Insights between Kidney and Brain," *Kidney Blood Press. Res.*, vol. 47, no. 8, pp. 493–505, 2022.

[136] K. J. Jensen, G. Alpini, and S. Glaser, “Hepatic Nervous System and Neurobiology of the Liver,” *Compr. Physiol.*, vol. 3, no. 2, pp. 1–20, 2013.

[137] D. Drachmann *et al.*, “Towards enhanced understanding of idiopathic ketotic hypoglycemia: a literature review and introduction of the patient organization, Ketotic Hypoglycemia International,” *Orphanet J. Rare Dis.*, vol. 16, no. 1, pp. 1–11, 2021.

[138] A. Scholze, V. Jankowski, L. Henning, W. Haass, and J. Jankowski, “Phenylacetic Acid and Arterial Vascular Properties in Patients with Chronic Kidney Disease Stage 5 on Hemodialysis Therapy,” pp. 1–6, 2007.

[139] J. Jankowski *et al.*, “Increased plasma phenylacetic acid in patients with end-stage renal failure inhibits iNOS expression,” *J. Clin. Invest.*, vol. 112, no. 2, pp. 256–264, 2003.

[140] M. A. Cinelli, H. T. Do, G. P. Miley, and R. B. Silverman, “Inducible nitric oxide synthase: Regulation, structure, and inhibition,” *Med. Res. Rev.*, vol. 40, no. 1, pp. 158–189, 2020.

[141] S. A. George, S. Al-Rushaidan, I. Francis, and M. R. Narayanan Nampoory, “2,8-Dihydroxyadenine Nephropathy Identified As Cause of End-Stage Renal Disease After Renal Transplant,” *Exp. Clin. Transplant.*, vol. 15, no. 5, pp. 574–577, 2017.

[142] P. Rinaldo, J. J. O’Shea, R. D. Welch, and K. Tanaka, “Stable isotope dilution analysis of n-hexanoylglycine, 3-phenylpropionylglycine and suberylglycine in human urine using chemical ionization gas chromatography/mass spectrometry selected ion monitoring,” *Biomed. Environ. Mass Spectrom.*, vol. 18, no. 7, pp. 471–477, 1989.

[143] P. Divry, C. Vianey-Liaud, and J. Cotte, “Gas chromatography--mass spectrometry (GC--MS) diagnosis of two cases of medium chain acyl-CoA dehydrogenase deficiency.,” *J. Inherit. Metab. Dis.*, vol. 7 Suppl 1, pp. 44–47, 1984.

[144] P. M. Prieto Jimenez, E. Jun-Ihn, M. Matthews, T. Lollie, Y. Qu, and M. G. Martin, “An Unusual Case of Infantile Hepatic Steatosis Caused by Coconut-Based Infant Formula,” *JPGN Reports*, vol. 3, no. 4, p. e235, 2022.

[145] S. Zhou *et al.*, “Spermine alleviates acute liver injury by inhibiting liver-resident macrophage pro-inflammatory response through ATG5-dependent autophagy,” *Front. Immunol.*, vol. 9, no. MAY, pp. 1–12, 2018.

[146] K. Igarashi, S. Ueda, K. Yoshida, and K. Kashiwagi, “Polyamines in renal failure,” *Amino Acids*, vol. 31, no. 4, pp. 477–483, 2006.

[147] M. Amin *et al.*, “Polyamine biomarkers as indicators of human disease,” *Biomarkers*, vol. 26, no. 2, pp. 77–94, 2021.

[148] M. E. Gherghina, I. Peride, M. Tiglis, T. P. Neagu, A. Niculae, and I. A.

Checherita, “Uric Acid and Oxidative Stress—Relationship with Cardiovascular, Metabolic, and Renal Impairment,” *Int. J. Mol. Sci.*, vol. 23, no. 6, 2022.

[149] D. Jiao *et al.*, “Changes in aging-induced kidney dysfunction in mice based on a metabolomics analysis,” *Front. Endocrinol. (Lausanne).*, vol. 13, no. September, pp. 1–15, 2022.

[150] M. K. Mazumder, B. C. Phukan, A. Bhattacharjee, and A. Borah, “Disturbed purine nucleotide metabolism in chronic kidney disease is a risk factor for cognitive impairment,” *Med. Hypotheses*, vol. 111, no. December 2017, pp. 36–39, 2018.

[151] M. N. Wamser *et al.*, “Effect of hypoxanthine, antioxidants and allopurinol on cholinesterase activities in rats,” *J. Neural Transm.*, vol. 120, no. 9, pp. 1359–1367, 2013.