Microchemical Journal 213 (2025) 113737

Contents lists available at ScienceDirect

Microchemical Journal

journal homepage: www.elsevier.com/locate/microc

Untargeted metabolomics HRMS data processing using regions of interest and multivariate curve resolution approaches to unveil health-to-disease transition

Luisa Barreiros [a,b,*], Benedita Sampaio-Maia [c,d], Inês Soares Alencastre [c,e], Romà Tauler [f], Marcela A. Segundo [a]

[a] LAQV, REQUIMTE, Department of Chemical Sciences, Faculty of Pharmacy, University of Porto, Rua Jorge Viterbo Ferreira 228, 4050-313 Porto, Portugal
[b] ESS, Polytechnic of Porto, Rua Dr. António Bernardino de Almeida 400, 4200-072 Porto, Portugal
[c] i3S - Instituto de Investigação e Inovação em Saúde/INEB - Instituto Nacional de Engenharia Biomédica, University of Porto, Rua Alfredo Allen 208, 4200-135 Porto, Portugal
[d] Faculty of Dental Medicine, University of Porto, Rua Dr. Manuel Pereira da Silva, 4200-393 Porto, Portugal
[e] Faculty of Medicine, University of Porto/RISE-Laboratório Associado, Alameda Prof. Hernâni Monteiro, 4200-319 Porto, Portugal
[f] Institute of Environmental Assessment and Water Research (IDAEA-CSIC), Department of Environmental Chemistry, Jordi Girona 18-26, 08034 Barcelona, Spain

ARTICLE INFO

Keywords:
Untargeted analysis
High-resolution mass spectrometry
Data processing
Chemometrics
Biomarkers
Chronic kidney disease

ABSTRACT

Early diagnosis has the potential to prevent or minimize the progression of diseases. The discovery of biomarkers that permit to pinpoint the health-to-disease transition is thus of crucial importance to improve diagnostic efficiency and treatment. Over the last years, untargeted metabolomic analysis of biomatrices has become a valuable tool to identify biomarkers of health status. However, it is still a recent research field with many challenges to address, namely the implementation of data treatment strategies that allow to deal efficiently with the highly complex data sets generated, without losing accuracy or relevant information. This work pursued the application of the chemometrics method Regions of Interest-Multivariate Curve Resolution (ROIMCR) to the MS-based metabolomic profiling of plasma samples aiming at the identification of chronic kidney disease (CKD) biomarkers.

ROIMCR successfully resolved the plasma profiles of three groups of individuals: healthy controls, intermediate stage (pre-dialysis) and end-stage (in dialysis) CKD patients. Positive (MS1+) and negative (MS1−) data were processed simultaneously without requiring previous time alignment, resulting in time-effective analysis with increased metabolite coverage and identification. Multivariate analysis (PCA, PLS-DA, ASCA) revealed that samples were clustered according to health status and permitted to determine the most influential metabolites in differentiating groups. Metabolites identification evidenced recognized biomarkers of CKD, validating the proposed approach, and potential new indicators of disease onset and progression.

A new analytical workflow involving instrumental analysis by UHPLC-HRMS and data treatment by chemometrics method ROIMCR, followed by multivariate analysis, is available for plasma metabolomic profiling. This platform can be easily adapted to other biomatrices and diseases, and applied to profile the same individual along time, fostering the development of personalized medicine.

1. Introduction

In the last two decades, metabolomics has become an increasingly popular approach for biomarker discovery, disease diagnosis, and biochemical pathways elucidation, especially when applied to compare the metabolic profiles of different groups, such as healthy controls and patients with a specific disease [1–3]. Mass spectrometry (MS) is the most widely used analytical technique in metabolomics [2,4]. Two main analytical strategies can be followed: targeted and untargeted analysis [1,2]. The targeted approach aims at the accurate quantification of

* Corresponding author at: LAQV, REQUIMTE, Department of Chemical Sciences, Faculty of Pharmacy, University of Porto, Rua Jorge Viterbo Ferreira 228, 4050-313 Porto, Portugal.
E-mail address: lbarreiros@ff.up.pt (L. Barreiros).

https://doi.org/10.1016/j.microc.2025.113737
Received 13 January 2025; Received in revised form 31 March 2025; Accepted 21 April 2025
Available online 23 April 2025

specific molecules of interest, requiring corresponding standards. Untargeted analysis permits the comprehensive study of all the molecules in a system and may provide valuable insights for the identification of unknown compounds, including health status indicators. Targeted analysis is frequently performed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) using low-resolution mass spectrometers, such as triple quadrupole. Untargeted analysis, aiming at providing the widest possible metabolome coverage, is mostly implemented by liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS), using quadrupole time of flight (Q-TOF) or Orbitrap analyzers [1,2,5].

Data acquisition in untargeted metabolomics is generally performed in full scan MS (MS1), followed by either a data-dependent acquisition (DDA) or data-independent acquisition (DIA) approach. DDA mode selects only certain analytes (e.g., the most abundant) and then fragments them (MS2), whereas in the DIA mode, MS2 spectra are produced for all precursor ions in defined m/z ranges [1]. Experiments of untargeted metabolomics generate highly complex data sets that require extensive treatment and processing for their correct interpretation [4,6].

Data treatment strategies involve the steps of raw data acquisition, data storage and conversion, import, compression, normalization, scaling and transformation, feature detection or peak resolution, biomarker screening and identification, and biochemical interpretation. Data compression is one of the most challenging steps, since it must reduce the original dimensions of the data (gigabytes of storage) without significantly compromising the information contained within and avoiding any loss of spectral accuracy. As untargeted analysis is a recent and hot topic of research, new methods, software and platforms for data treatment have been developed during the last years [7–12]. Frequently, the post-experimental data processing is performed by commercial software provided by the MS equipment manufacturers [3,13], using fixed or not disclosed methodologies, being thus difficult to exactly know the algorithm description and workflow. In alternative, chemometric methods may be used for implementation of data processing strategies that are more comprehensive and tailored to each metabolomics study [14]. Recently, the chemometric method Regions of Interest-Multivariate Curve Resolution (ROIMCR) has been applied successfully in metabolomics studies [12,15–23]. The ROI procedure has been demonstrated more advantageous than the classical binning during the compression step as no loss of spectral accuracy occurs [15,16]. In turn, the MCR-Alternating Least Squares (MCR-ALS) methodology is a powerful tool that, after data compression, allows the resolution of coeluting compounds and the retrieval of the pure spectra and elution profiles of the most relevant metabolites present in a sample, without requiring peak alignment or shaping corrections, since chromatograms are only matched in the mass spectral direction [15,24–28]. ROIMCR has been used in environmental and metabolomic studies, to evaluate the toxic effects of pharmaceutical compounds and chemical contaminants on different aquatic organisms [12,19,22,23,29], and also in the untargeted protein analysis of samples from wastewater treatment plants (WWTP) [21]. On the other hand, the application of ROIMCR to human biological samples has been only described in a single study pursuing the analysis of polar compounds and the identification of prostate cancer biomarkers in urine samples [18].

Having this in mind, this work pursued the untargeted metabolomic profiling of human plasma samples using LC-HRMS for instrumental analysis and the chemometrics method ROIMCR for data treatment, in the search of potential biomarkers of health-to-disease transition. In fact, even with optimized chromatographic conditions, strong coelution of compounds occur in the analysis of complex samples like plasma, hampering the direct interpretation of the acquired data and demanding for the application of adequate post-run data processing strategies. Chronic kidney disease (CKD) was selected as case study as it represents a global health problem [30,31] and, despite the identification of several potential biomarkers [32–36], finding metabolites that can serve as indicators of disease onset and progression remains challenging and of utmost importance. CKD includes heterogenous disorders that affect progressively the structure and function of the kidney. The kidney function is commonly assessed by the estimated glomerular filtration rate (eGFR), calculated using serum creatinine, age, sex, and race of the patient. However, this currently used biomarker is limited and lacks sensitivity as it is influenced by individual characteristics, and variations are only detected when the renal function has already been compromised [36]. To our knowledge, this is the first time ROIMCR is applied to the analysis of plasma samples and to CKD research.

2. Materials and methods

2.1. Chemicals and solutions

Ultrapure water was produced in an Arium water purification system (resistivity $>$ 18 MΩ cm, Sartorius, Göttingen, Germany) and used for the preparation of all solutions. Acetonitrile (LiChrosolv LC-MS grade) and formic acid were supplied by Merck (Darmstadt, Germany). The aqueous component of mobile phase was 0.1 % (v/v) formic acid in water whereas 0.1 % (v/v) formic acid in acetonitrile was the organic counterpart. Both solutions were previously filtered through Millipore filters (Billerica, MA, USA) and degassed in an ultrasonic bath for 15 min prior to use.

2.2. Clinical samples and compliance with ethical standards

This study included 24 Caucasian adult participants that were divided in three groups: group A: 8 healthy control individuals (living kidney donors), group B: 8 pre-dialysis CKD patients (intermediate stage), and group C: 8 end-stage kidney disease (ESKD) patients undergoing peritoneal dialysis. As this work aimed at validating if the ROIMCR method could be successfully applied to the MS-based metabolomic profiling of human plasma samples towards the identification of CKD biomarkers, including the comparison with previously determined potential biomarkers and the discovery of new indicators, eight samples per group were considered sufficient to demonstrate the usefulness of the proposed methodology.

The protocol of the study was authorized by the local Ethics Committee (*Comissão de Ética para a Saúde do Centro Hospitalar Universitário de São João (CHUSJ)*, ref. CES 87-15 and 130/19) and conducted in agreement with the 1964 Helsinki declaration and subsequent revisions. All participants were followed up in the Nephrology Department of *CHUSJ* and gave their free written informed consent before sample collection. The clinical and demographic information of each participant were also registered. The characterization of the three study groups including participants' age, sex, eGFR and CKD staging can be found in Supplementary data (Table S1).

2.3. Sample treatment

A blood sample was collected from each participant, followed by centrifugation at 3892*g* for 15 min at 4 °C. The resulting plasma was immediately frozen and stored at −80 °C. Prior to LC-MS analysis, plasma samples (100 μL) were thawed at room temperature and treated with 300 μL of acetonitrile containing 0.1 % (v/v) formic acid to precipitate proteins. After vortexing during 60 s, treated samples were maintained at 4 °C for 10 min to enhance protein precipitation. Following this period, a centrifugation step at 18,000*g* for 8 min at 4 °C was implemented. Supernatants were collected, evaporated to dryness under a gentle stream of nitrogen, and finally reconstituted in 100 μL of acetonitrile–water (5:95, v/v) containing 0.1 % (v/v) formic acid for instrumental analysis. This solvent composition ensured the solubilization of plasma metabolites and matched closely the mobile phase composition at the beginning of the chromatographic run, preventing analytes pre-elution and peak broadening that could compromise the resolution of the analysis.

2.4. Instrumental analysis

2.4.1. Ultrahigh-performance liquid chromatography coupled to high resolution mass spectrometry detection (UHPLC-HRMS) analysis

For metabolomic profiling of plasma samples, each sample extract (5 μL) was separated on a UHPLC-HRMS Vanquish equipment (Thermo Fischer Scientific, Bremen, Germany) using a reversed-phase Zorbax Eclipse Plus RRHD C18 column (100 mm × 2.1 mm, 1.8 μm particle size, Agilent, California, USA). Elution was implemented in gradient mode with 0.1 % (v/v) formic acid in water as solvent A and 0.1 % (v/v) formic acid in acetonitrile as solvent B, at a total flow rate of 0.3 mL min^{-1}. The selected gradient started with 1 % of solvent B, which was maintained during 1 min and then increased to 40 % within 5 min, 50 % within 8 min, 65 % within 10 min and 76 % within 16 min. From 16 to 20 min, the contribution of solvent B was changed to 100 %. This composition was maintained up to 25 min, after which the initial conditions were re-established at 27 min and held during 3 min more to assure column equilibration. Hence, the total run time was 30 min. The column temperature was 40 °C.

HRMS analyses were done on an Orbitrap Exploris™ 120 mass spectrometer (Thermo Fischer Scientific) controlled by Orbitrap Exploris™ Tune Application 2.0.185.35 and Xcalibur 4.4.16.14 software. The capillary voltage of the electrospray ionization source (ESI) was set to 3.5 kV. The capillary temperature was 300 °C. The sheath gas and auxiliary gas flow rates were at 50 and 10 (arbitrary units as provided by the software settings). Data were acquired in full scan mode (MS1) over the *m/z* range 100–1000, at a resolution of 30,000. Data acquisition was performed simultaneously in positive and negative ionization modes, i.e., in the same chromatographic run. Moreover, data-dependent acquisition (DDA) MS/MS (MS2) fragmentation was performed on high-energy collisional dissociation (HCD) using nitrogen as collision gas with collision energy settings of 60 %, at a resolution of 15,000. Precursor ions for MS2 analysis were selected using the DDA Top4 method, i.e., the 4 most intense ions in each MS1 scan. The AGC target was set to standard mode and the maximum injection time was set to auto for both MS1 and MS2 analyses.

2.4.2. Analytical run order and quality controls

To promote the acquisition of high-quality data and the subsequent success of metabolomics analysis, particular attention was given to the analytical sequence order and to the inclusion of quality controls (QC) [37]. QCs corresponded to pooled samples and permitted to assure that the detected variation was not attributed to instrumentation but to features of the samples. The order in which the study samples were analyzed and the frequency of QC samples injections were designed considering different factors, namely the complexity of plasma matrix, total number of samples, study groups, batch length, system specifications, and data processing software, towards ensuring the quality of the acquired data. The analytical run order is presented in detail in Supplementary data (Table S2). The run started with multiple injections of solvent blanks (acetonitrile–water (5:95, v/v)) and pooled QC samples to condition the analytical system, followed by a continuous sequence of extraction blanks, samples in random order in sets of 9 and QCs (Table S2). A solvent blank was injected at the end of the run as a cleaning procedure. Extraction blanks were obtained using 100 μL water (instead of plasma) that was submitted to the same sample treatment steps: addition of organic solvent, vortexing, refrigeration, centrifugation, supernatant evaporation to dryness, and reconstitution in acetonitrile–water (5:95, v/v) containing 0.1 % (v/v) formic acid (see Section 2.3. Sample treatment). Five types of QCs were prepared and analyzed. QC_ALL was prepared by mixing equal parts of all the plasma samples. QC_A, QC_B and QC_C corresponded to mixtures of equal parts of the samples from each group included in the study (A, B and C). Finally, QC_SET was prepared by mixing the set of samples analyzed in each part of the sequence. Each QC was divided in aliquots to perform 4 (or 5 in the case of QC_ALL) independent extractions of each pooled QC sample.

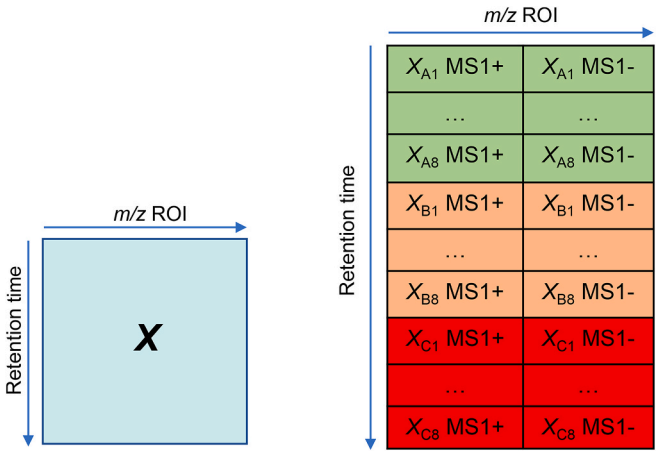

Fig. 1. Data matrices obtained after ROI analysis of a single sample (left) or the subset of all samples (right).

QC samples were processed following the same procedure as study samples and extraction blanks. Additionally, extraction replicates and instrumental replicates were also included for one sample of each group of individuals (see Table S2).

The data from the solvent blanks and pooled QC samples used to condition the analytical system were not considered for the subsequent post-experimental data processing.

2.5. Data processing

2.5.1. Regions of interest (ROI) procedure

In a first step, the raw datasets obtained from instrumental analysis (.raw files) were converted to mzXML format using the MSConvert tool from ProteoWizard open-source software [38]. A "peak peaking" filter was applied to only extract and convert the MS1 data. The generated mzXML files were then imported to the MATLAB computational environment (release R2020b, The MathWorks, Inc., MA, USA) for implementation of data processing procedures, using its Bioinformatic toolbox and the MSroi GUI tool as previously described [16]. The data from the samples of each group (A, B, C) and of each type of QC were imported simultaneously, using the multiple runs option, resulting in a column-wise augmented matrix.

Before ROI analysis, the data obtained in positive (MS1+) and negative (MS1−) ionization modes were separated. Moreover, only the data acquired until 18.5 min (1110 s) of each chromatographic run were considered for further data processing as the remaining gradient program corresponded to column wash and equilibration.

The ROI procedure was applied for data compression and filtering, towards selecting only the reliable and interesting MS signals (MS ROIs) for subsequent analysis, thus reducing the size of data sets and computer storage, without losing mass resolution or chemical information. The selection of MS ROIs is performed according to three parameters: intensity threshold, mass error and minimum number of occurrences to define a peak [12,15,16]. The optimal value of each parameter must be tested in each case as it is influenced by the resolution and the scan speed of the used mass spectrometer [12,14]. As a result, a data matrix is obtained for each sample, with the different elution/retention time values in its rows and the selected *m/z* ROI values in its columns (Fig. 1). In this work, ROI analysis was performed (i) separately for each group of samples and QCs, (ii) considering the groups two by two (A versus B, A versus C, B versus C), and (iii) considering all groups and QCs together. The ROI analysis of MS1+ and MS1− datasets was performed independently. No time alignments were required for the simultaneous analysis of the different datasets of the same ionization mode as the vertically augmented data matrix is formed by column concatenation of the individual ROI data matrices. Before proceeding, the MS ROIs

exhibiting low resolution and detected in blank samples, probably corresponding to solvent artifacts, were removed from the data matrices to minimize interferences and improve the quality of analysis. Finally, the column-wise augmented data matrix was augmented horizontally, i.e., row-wise augmented to include the data obtained in positive and negative mode for each sample (Fig. 1). No previous time alignments were required before the horizontal concatenation of data matrices as MS1+ and MS1− data were acquired in the same run and with matching retention times.

2.5.2. Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) method

The next step of data processing consisted in applying the MCR-ALS method to the data matrices (***X***) resulting from ROI procedure. MCR-ALS [4,24,25,27] performs the bilinear decomposition of the data matrix ***X*** into the product of two factor matrices, ***C*** and $\boldsymbol{S}^T$ (eq 1). The matrix ***C*** contains the elution profiles of *N* components (sample constituents) whereas the matrix $\boldsymbol{S}^T$ contains the corresponding mass spectra. The part of matrix ***X*** that is not explained by the model forms the residual matrix ***E***.

$$X = C \times S^T + E \quad (1)$$

MCR-ALS analysis permits to explain the data matrix ***X*** variance through a reduced number of components characterized by their elution and spectra profiles. The number of components must be sufficient to explain the majority of data variance arising from MS signals of sample constituents. An initial estimation of the number of components is obtained through singular value decomposition (SVD) of the data matrix ***X*** [24]. Moreover, a preliminary estimation of the associated elution and spectra profiles is mathematically derived from the purest variables (*m/z* values) or samples (times) using methods such as SIMPLISMA [39]. An alternating least squares algorithm is then applied to optimize elution (***C***) and spectral ($\boldsymbol{S}^T$) profiles iteratively until convergence and optimal fitting of the experimental data matrix ***X*** are attained. This optimization is performed under some constraints that in this work included non-negativity and normalization of components' spectra to the maximum height equal to one (spectra equal height).

The resolved mass spectra in $\boldsymbol{S}^T$ matrix contain the ion signals which can be subsequently used for identification of chemical compounds present in the analyzed samples. In the present study, the spectra obtained in positive mode (MS1+) and in negative mode (MS1−) are simultaneously displayed for each resolved component, which widens metabolite coverage and benefits the identification process. On the other hand, the resolved elution profiles in ***C*** matrix can be associated with the concentration of sample constituents. The peak areas and the peak heights of the resolved elution profiles can be used for the relative quantification of each resolved component (sample constituent) in the different analyzed samples. A more detailed description of MCR-ALS analysis can be found in previous reports [12,15,23–25].

The MCR-ALS method was applied to resolve the elution and spectral profiles of the ROI data matrices resulting from (i) the independent analysis of each group of samples and QCs, (ii) the analysis of the different groups considered two by two (A versus B, A versus C, B versus C), and (iii) the simultaneous analysis of all groups of samples and QCs. The peak heights of the MCR-ALS resolved elution profiles were used for further analysis by statistical tools and to determine the metabolites varying between groups. The resolved mass spectra and the highest intensity *m/z* values in each resolved component were used to identify those metabolites.

2.5.3. Statistical analysis

The matrix with the peak height values obtained for every resolved component (sample constituent) in the different samples and QCs was used to investigate the differences in the metabolic profile of the groups of samples. To accomplish this goal, the peak heights matrix was analyzed by Principal Component Analysis (PCA) [40], Partial Least Squares-Discriminant Analysis (PLS-DA) [41] and Analysis of Variance (ANOVA)-Simultaneous Component Analysis (ASCA) [42], using the PLS_Toolbox 9.0 (Eigenvector Research, Inc., WA, USA). The theoretical description of each multivariate analysis method can be found in Supplementary data. Statistical analysis by PCA, PLS-DA and ASCA was performed (i) considering the total number of components (sample constituents) resolved by ROIMCR data processing and (ii) removing the ROIMCR components with a poorly defined profile, which could not be ascertain to chemical compounds of the samples.

PCA models were constructed using autoscaling preprocessing and the SVD algorithm. Cross-validation was performed by the venetian blinds method and the two principal components (PC) explaining the most variance were finally selected. First, the peak heights values resulting from the MCR-ALS processing of the augmented ROI data matrix containing all the samples (groups A, B and C) and QC controls were analyzed by PCA to have an overview on the patterns and relations between samples and controls (clustering). For a more detailed analysis on the distribution of samples, PCA was also applied to the peak heights matrices resulting from the comparisons between two groups of samples (A versus B, A versus C, B versus C).

PLS-DA analysis was subsequently applied to build regression models and further discriminate the differences between two groups of samples. A class vector was used to identify the two groups to be compared in each analysis and then regressed against the peak heights of ROIMCR components. Autoscaling preprocessing was used for both X-block and Y-block data. Cross-validation of the constructed model was performed by the venetian blinds approach and the number of latent variables (LV) finally selected was two. From the outputs of PLS-DA analysis, it was possible to determine the variables (resolved components/metabolites) contributing most for the separation between sample groups. Components presenting Variable Importance in Projection (VIP) scores > 1 were selected as the most influential metabolites, i.e., responsible for the differentiation between samples [18,43]. These metabolites were considered potential biomarkers and were subsequently identified (see section 2.6).

Finally, ASCA was applied to the same peak heights matrices to evaluate the statistical significance of the differences observed between groups. ASCA was implemented without preprocessing and using a limit of 10,000 permutations to statistically assess the effects induced by the health condition factor.

2.6. Identification of metabolites

To identify the metabolites varying significantly between groups, the most intense *m/z* values in each significant MCR-ALS resolved component (MS1+ and MS1−) were searched in existing public databases of human metabolites, particularly the Human Metabolome Database (HMDB) (https://hmdb.ca/) [44,45]. This search was performed according to three predefined criteria: (i) *m/z* value to search introduced with 4 decimal places, (ii) mass error value (tolerance) $\leq$ 5 ppm, (iii) adduct type limited to $[M + H]^+$, $[M + Na]^+$, $[M + ACN + H]^+$, $[M + ACN + Na]^+$, $[M + H - H_2O]^+$, and $[M + H + Na]^{2+}$ in positive ionization mode and $[M-H]^-$, $[M-H_2O-H]^-$, $[M + FA - H]^-$, $[M + Na - 2H]^-$, $[M-2H]^{2-}$, and $[2M-H]^-$ in negative ionization mode. The identification obtained in HMDB was then confirmed with the MS2 data obtained by data dependent analysis (DDA). The comparison of the fragmentation pattern of the identified metabolite compiled in databases with the pattern experimentally obtained was performed using the Competitive Fragmentation Modeling for Metabolite Identification online tool (CFM-ID) (https://cfmid.wishartlab.com/) [46,47] and MassBank (https://massbank.eu/).

The obtained metabolite identification corresponds to confidence level 2 according to the Metabolomics Standards Initiative [48] or confidence level 3 according to Schymanski et al. [49], as evidence exists for a possible structure, MS/MS experimental data is available and

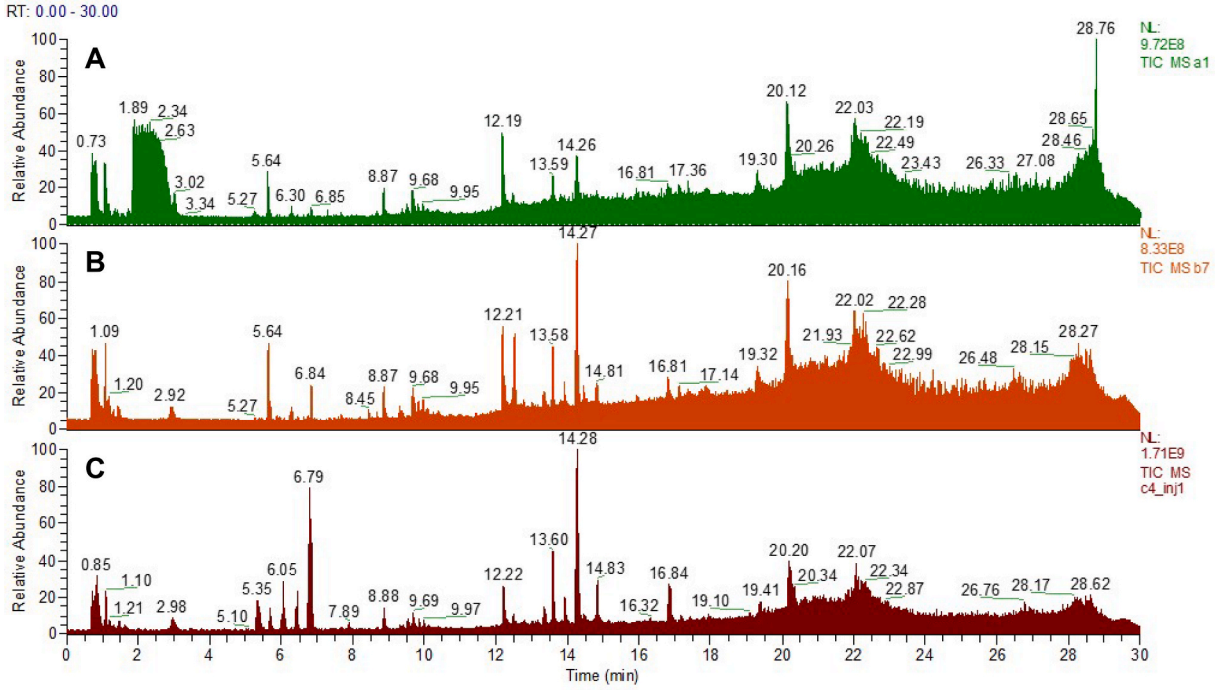

Fig. 2. Representative total ion current (TIC) chromatograms of plasma samples from the different groups under study: healthy controls (A), pre-dialysis CKD patients (B), and CKD patients at final stage (ESKD), undergoing peritoneal dialysis (C).

has been compared with in silico fragmentation of candidates, but the analysis and comparison with a reference standard has not been conducted yet.

3. Results and discussion

3.1. LC-MS analysis of plasma samples

Sample preparation for untargeted analysis should assure metabolome stability while preventing losses of metabolites and the presence of artifacts [2]. The ideal sample preparation method should combine comprehensive extraction of metabolites, simplicity, quickness, easy implementation with a minimum number of steps, and reproducibility [2,5]. Hence, protein precipitation through the addition of an organic solvent (acetonitrile) was selected to treat plasma samples before LC-MS analysis, aiming at simple and minimal sample treatment and at increased metabolite coverage. Chromatographic separation was implemented using a modified reversed-phase C18 column and gradient elution with a slow increase of organic modifier content from 1 to 76 % to profile both hydrophilic and hydrophobic compounds and maximize analytes separation. An example of the total ion current (TIC) chromatograms obtained for the different groups of individuals under study (A, B, and C) is presented in Fig. 2. The observation of the elution patterns evidenced the existence of differences between groups. The initial part of the chromatograms (0–4 min), corresponding to the elution of more hydrophilic compounds, presented higher intensity and higher diversity of peaks in the case of healthy controls (group A). On the other hand, mid-polar and non-polar compounds, i.e., hydrophobic compounds eluting with higher organic modifier percentages (retention time > 4 min) were more abundant in disease patients (groups B and C). After 10 min, the chromatograms exhibited high density of compounds that required advanced data processing and treatment by the ROIMCR chemometrics methodology.

3.2. ROIMCR analysis of study samples

In ROI analysis, the intensity threshold finally selected to eliminate background and instrumental noise without losing significant information was 1×10^6, which is within the recommended 0.1–1 % of the maximum MS intensity value [15]. The mass error tolerance was defined to 0.005 Da in accordance with the resolution of the used Orbitrap mass analyzer. Finally, the minimum number of MS signals to define a chromatographic peak was adjusted to (i) 5 in the case of the ROI analysis performed separately for each group of samples and QCs, and to (ii) 20 when considering the groups two by two or all groups and QCs together. The values for the three parameters were equal for the ROI analysis of positive (MS1+) and negative (MS1−) ionization datasets. After removal of solvent artifacts, the resulting number of ROIs varied between 234 and 660 for data matrices obtained for each individual group in positive mode ($\boldsymbol{X_{MS1+}}$) and between 212 and 487 for the negative

Table 1
Results of ROI analysis of samples from groups A, B and C and QCs.

Analysis	X_{MS1+}		X_{MS1-}		X_{MS1+-}	
	Nr. of retention times	Nr. of ROIs	Nr. of retention times	Nr. of ROIs	Nr. of retention times	Nr. of ROIs
Group A	10,640	481	10,640	299	10,640	780
Group B	10,648	451	10,648	304	10,648	755
Group C	10,680	660	10,680	487	10,680	1147
QC_A	5328	599	5328	363	5328	962
QC_B	5316	234	5316	212	5316	446
QC_C	5340	380	5340	329	5340	709
QC_ALL	5336	342	5336	260	5336	602
A versus B	21,288	548	21,288	338	21,288	886
A versus C	21,320	764	21,320	529	21,320	1293
B versus C	21,328	705	21,328	521	21,328	1226
A + B + C + QC_ALL	37,304	796	37,304	557	37,304	1353
A + B + C + QCs	53,312	852	53,312	600	53,312	1452

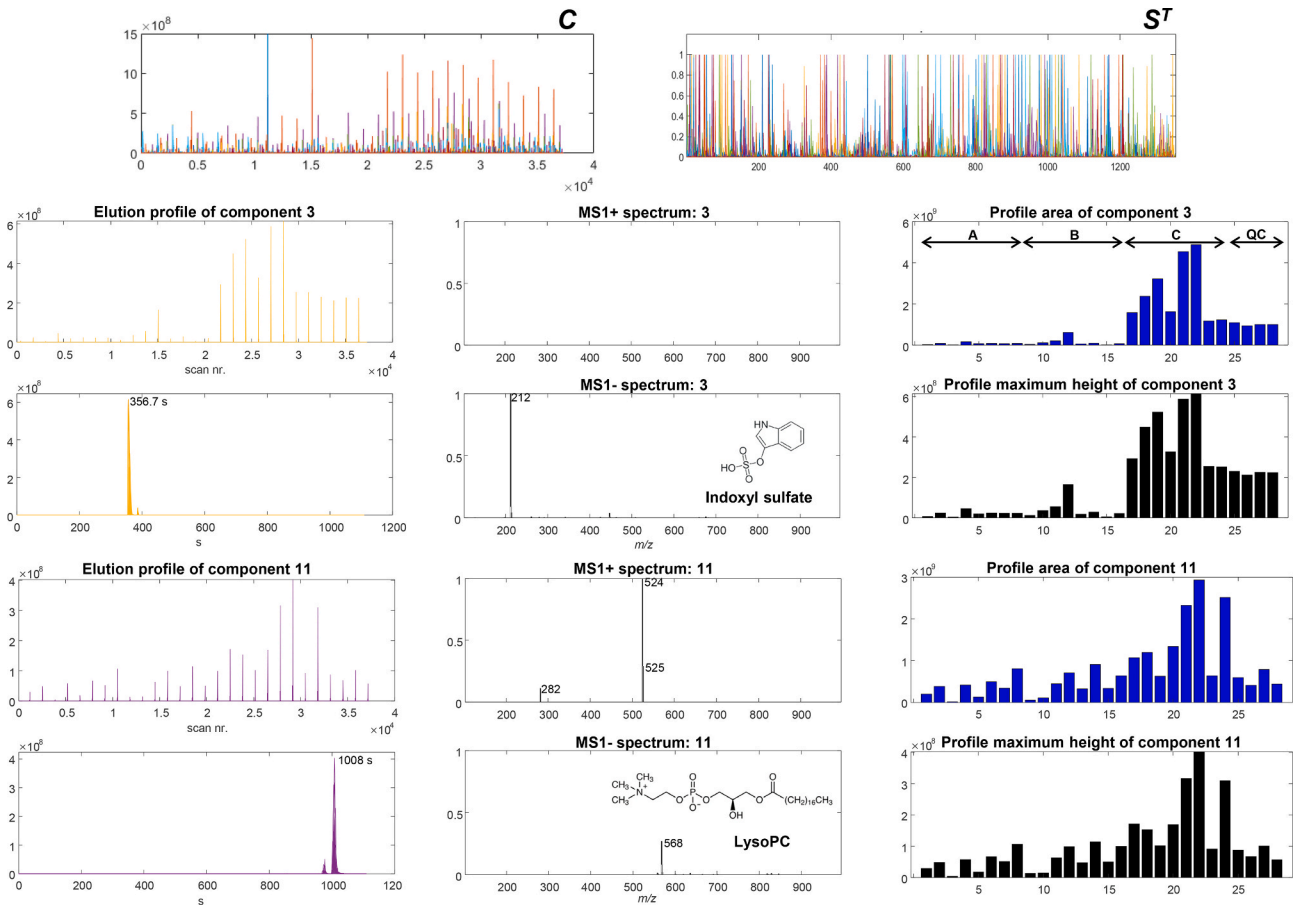

Fig. 3. MCR-ALS results in the analysis of the ROI data matrix resulting from the simultaneous analysis of all groups of samples and QC_ALL. The two plots at the upper part represent the elution ($\boldsymbol{C}$) and spectra ($\boldsymbol{S}^T$) profiles of all the resolved components. The plots at the lower part represent the separate results obtained for components 3 and 11. Left panel: resolved elution profiles in all samples and controls displayed side by side (top) and overlaid (bottom). Center panel: MS1 spectrum obtained in positive (top) and negative (bottom) ionization modes. Right panel: peak area (top) and peak height (bottom) values obtained from the elution profiles resolved for the different samples and controls.

mode ($\boldsymbol{X}_{MS1-}$) counterparts (Table 1). The number of ROIs was higher in the group of CKD patients at final stage (ESKD) (group C), which is in accordance with the higher abundance of peaks depicted in the corresponding chromatograms (Fig. 2). The combined analysis of the different groups of samples considered two by two and the analysis of all samples and QCs together yielded $\boldsymbol{X}_{MS1+}$ and $\boldsymbol{X}_{MS1-}$ data matrices with 548–852 and 338–600 ROIs, respectively (Table 1). In all cases, the number of ROIs obtained in MS1− matrices was lower than the number of ROIs obtained in MS1+ matrices which is expected considering the use of electrospray ionization and the higher number of compounds ionizing in positive mode in comparison to negative mode [2,50]. The augmented data matrices obtained by horizontal concatenation of positive and negative mode data ($\boldsymbol{X}_{MS1+-}$) presented a total number of ROIs within the range 446–1452, corresponding to the sum of ROIs obtained in the data matrices of each ionization mode (Table 1). The number of retention time values increased with the number of samples simultaneously analyzed due to the vertical concatenation of the different datasets considered. Hence, the number of retention time values ranged from 5316 to 10,680 in the data matrices of each group, from 21,288 to 21,328 in the comparison of two groups, and from 37,304 to 53,312 in the super-augmented data matrices comprising all samples and QCs (Table 1).

After ROI analysis, MCR-ALS was subsequently applied to the augmented data matrices ($\boldsymbol{X}_{MS1+-}$) resulting from the independent analysis of each group of samples, the analysis of the different groups considered two by two, and the simultaneous analysis of all samples and QCs. The number of components initially proposed in MCR-ALS analysis was 100, which permitted to explain a high percentage of the data variance, between 98.90 and 99.69 % of fit (Table S3). After careful evaluation of the profiles of the 100 components resolved in the MCR-ALS analysis of the different data matrices, the components presenting poor-defined elution profiles, unreliable mass spectra and explaining little data variance were removed for further analysis since they could not be assigned to sample constituents (i.e., they correspond to background, solvent, instrumental noise…). The number of components finally resolved and selected was between 86 and 92 and still explained ≥95.24 % of the observed data variance (Table S3).

An example of the output of MCR-ALS analysis is displayed in Fig. 3 for the augmented data matrix containing all samples and QC_ALL. The global results obtained for resolved elution ($\boldsymbol{C}$) and spectral profiles ($\boldsymbol{S}^T$) are presented in the upper part of the figure, evidencing good chromatographic and spectral features but being difficult to distinguish among them. Thus, the results obtained for every component (sample constituent) were then plotted individually to visualize their elution profiles in the different samples simultaneously analyzed, as well as their corresponding mass spectra obtained in positive and in negative modes, and their elution peak area and height values in every analyzed sample. As an example, the results obtained for components 3 and 11 are presented in Fig. 3. In the left panel, the elution profiles of the two components in the different samples and QC_ALL are displayed side by side (top) or overlaid (bottom). The overlaid representation permitted to verify that samples and QCs exhibited matching retention times for the two components, indicating the adequacy of data processing procedures and the correct resolution of components. The elution profiles were

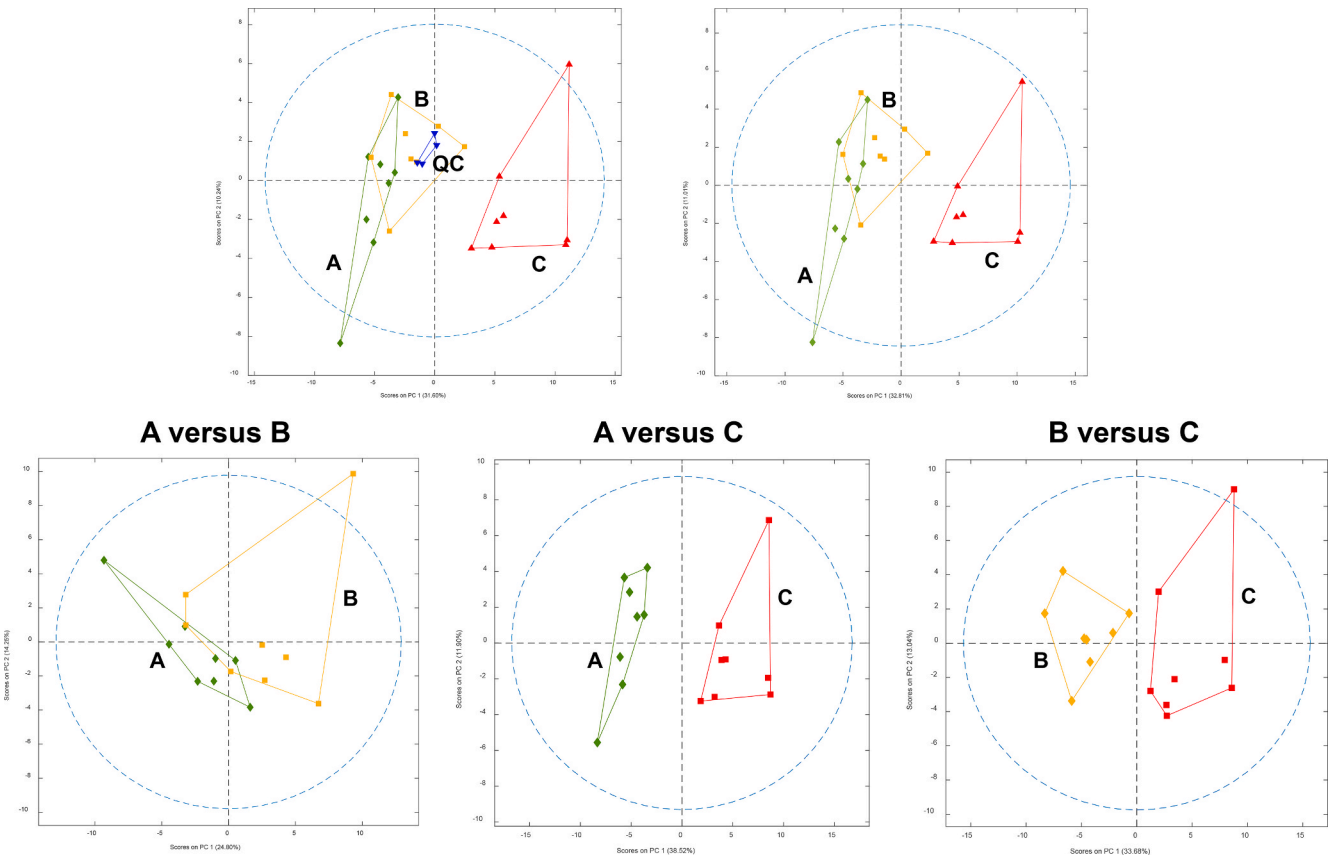

Fig. 4. PCA scores plot resulting from the analysis of peak height values of the ROIMCR resolved elution profiles of all the samples and QC_ALL (top panel, left), all the samples without QCs (top panel, right), and comparison of the different groups considered two by two (bottom panel, A versus B, A versus C, B versus C). A, healthy controls; B, pre-dialysis CKD patients; C, CKD patients at final stage (ESKD), undergoing peritoneal dialysis; QC, quality control prepared by mixing equal parts of all the plasma samples under study (QC_ALL).

directly associated with their mass spectra, linking the positive (MS1+, top) and negative (MS1-, bottom) ionization modes spectra obtained for each component (center panel). The highest intensity *m/z* values were used to identify those metabolites. Components 3 and 11 were later identified as indoxyl sulfate and a lysophosphatidylcholine (LysoPC (18:0/0:0)), respectively (Tables S5 and S6). Component 3 only ionized in negative mode forming the anion $[M-H]^-$ at *m/z* 212.0019 that corresponded to deprotonated indoxyl sulfate. In turn, component 11 (LysoPC(18:0/0:0)) was ionized in both modes, forming the $[M + H]^+$ cation detected in MS1+ spectrum at *m/z* 524.3688 and the negatively charged adduct ion with formic acid $[M + FA - H]^-$ detected in MS1- spectrum at *m/z* 568.3609 (Fig. 3, center panel). These results evidenced the benefits of simultaneously analyzing positive and negative ionization data and linking the corresponding MS1+ and MS1− spectra. In fact, if only positive ionization mode was employed, several relevant metabolites such as indoxyl sulfate could not be detected. Moreover, the identification process in the case of metabolites ionizing in both modes is facilitated and strengthened, as observed for component 11 corresponding to a lysophosphatidylcholine. The peak area (top) and peak height (bottom) values displayed in the right panel of Fig. 3 were used for the relative quantification of the resolved metabolites in the different analyzed samples. Both metabolites presented increased levels in disease patients, which is in accordance with previous studies reporting the accumulation of uremic toxins such as indoxyl sulfate in CKD patients and the dysfunction of lipid metabolism [35,36,51,52]. For more details, please see Section 2.6, for the identification of metabolites, and Section 3.4, for the identification of potential biomarkers of CKD.

To our knowledge, the ROIMCR method has not been applied for metabolomic profiling of human plasma samples using simultaneously the data acquired in positive and negative ionization modes, which represents an improvement compared to previous works in this field and progress beyond the state of the art. In previous studies, acquisitions in positive and negative mode were performed in different chromatographic runs, requiring the previous alignment of retention times between the two data sets before horizontal concatenation and analysis, and thus resulting in more time-consuming data processing schemes [23,53]. Moreover, this work reports the results of ROIMCR for the analysis of untargeted metabolomics experiments involving CKD. Previous works dedicated to the study of metabolomic profiles in patients with CKD by untargeted analysis have used commercial software supplied by MS instrument manufacturers [33,34]. Most of these works have been based on data acquired in a single ionization mode, mainly in positive mode, or if both ionization modes were used, data were acquired in different runs and data processing was performed independently, reinforcing the novelty of the present work.

3.3. Comparison of plasma metabolic profiles by multivariate statistical analysis

Considering that peak height values of every component (sample constituents or metabolites) finally resolved by MCR-ALS reveal changes in their concentrations across the different samples, these values were further analyzed to differentiate samples according to their health condition and to investigate their patterns. PCA was first applied to the analysis of all samples and QCs, considering the number of components (sample constituents) finally resolved by ROIMCR, after removal of those components with a poorly defined elution profile ($N = 90$ components, Table S3). The obtained scores plot (Fig. 4, left) evidenced that

PC1 and PC2 explained, respectively, 31.6 and 10.2 % of the observed variance, representing a total explained variance of 41.8 %. Samples were mainly distributed across PC1 and formed three groups, although there was a partial overlapping between samples from group A (healthy controls) and group B (pre-dialysis CKD patients) (Fig. 4, top panel, left). Samples from group A (healthy controls) were clustered more closely to samples from group B (pre-dialysis CKD patients) than samples of group C (CKD patients at final stage, ESKD), which were more distinctly separated. The three groups exhibited some degree of dispersion, attributed to the analysis of a very complex sample biomatrix, coming from individuals presenting different characteristics. Furthermore, the QC_ALL samples, corresponding to the mixture of all the samples, were clustered near the center of the PCA scores plot, confirming the adequacy of both the implemented experimental and the data analysis procedures. PCA was also performed considering only the study samples, i.e. without QCs, and the same trend was observed (Fig. 4, top panel, right). In this case, the total explained variance was 43.8 % with PC1 and PC2 contributing 32.8 and 11.0 %, respectively. For a more detailed analysis of the patterns and relations between samples, three more PCA models were created, comparing the three groups of samples two by two and considering the number of ROIMCR components finally selected (A versus B, $N = 86$ components; A versus C, and B versus C, $N = 91$ components, Table S3). The output of these analyzes (Fig. 4, bottom panel) confirmed that the differences and groups could be distinguished, with PC1 explaining most of the variance between the plasma samples (24.8–38.5 %), and with a total explained variance ranging from 39.0 to 50.4 % for PC1 + PC2. Healthy individuals were separated from CKD patients at final stage (A versus C) and CKD patients at an intermediate stage of disease (A versus B), although partially overlapped in the latter case. Furthermore, the two groups of CKD patients (B versus C) were clearly distinguished (Fig. 4, bottom panel). An additional PCA model was created considering all the study samples, QC_ALL and the QCs prepared by mixing the samples of each group (QC_A, QC_B and QC_C) and including $N = 90$ components finally resolved by ROIMCR (Table S3 and Fig. S1). The QCs for each group of samples clustered close to the center of the corresponding samples (A, B and C), evidencing that QCs share the variance within each specific group (Fig. S1).

PCA allows to explore the patterns and differences of groups but does not permit to distribute the samples in classes. This can be achieved by classification models such as PLS-DA that was also applied in this work. PLS-DA analysis was performed using the peak height values resulting from the comparison of two groups of samples (Fig. S2). Results confirmed that the different groups could be indeed discriminated from each other, in accordance with health status and the stage of disease. The percentage of variance explained by the two first latent variables (LV1 and LV2) was of the same order of magnitude compared to PCA components (PC1 and PC2): 21.3–38.3 % for PLS-DA LV1 and 9.7–13.2 % for PLS-DA LV2 (Fig. S2, top). Overall, PLS-DA models using the two first latent variables could explain between 34.4 % (A versus B) and 48.0 % (A versus C) of the X-variance (peak heights), and between 94.3 % (A versus B) and 98.6 % (A versus C) of the Y-variance (class, health status). These results imply that the model could effectively explain the class separation and the discrimination between study groups. They also evidenced that the X-variance is partially correlated with the Y-variance, indicating that only some of the sample constituents can be used as markers to detect health changes. Given the complexity of plasma samples and the high number of metabolites detected, this observation was expected. PLS-DA VIP scores permitted to determine the most influential metabolites whose concentration changes enabled the differentiation between groups (Fig. S2, bottom). The metabolites associated with the ROIMCR components whose peak heights exhibited VIP scores above 1 were considered responsible for the health variability among samples and, consequently, potentially serving as biomarkers. In the comparison A versus B, a total of 16 ROIMCR components (metabolites) presented VIP > 1, whereas a higher number of 46 and 39 components yielded VIP > 1 in the comparisons A versus C and B versus C, respectively (Fig. S2, bottom). These were the metabolites and potential biomarkers finally selected for subsequent identification by their m/z values.

The application of ASCA to the same peak heights matrices allowed to confirm statistically significant ($p < 0.005$) effects on health conditions of the metabolite concentration differences observed between groups. Additionally, paired t-test evidenced that those components (metabolites) presenting statistically significant differences ($p < 0.05$) were coincident with the ones presenting VIP scores > 1, confirming what were the most relevant metabolites for discrimination among different sample groups.

Statistical analysis (PCA, PLS-DA, ASCA) was also performed taking into account the total number of components simultaneously resolved by ROIMCR, i.e. 100 components. The multivariate analysis of the corresponding peak height values revealed the same clustering and separation between samples, but with lower percentages of explained variance (Figs. S1, S3 and S4). This observation confirmed the adequacy of removing the components with poorly defined profiles, resulting in better separation among sample groups.

3.4. Identification of potential biomarkers of CKD

The identification of the metabolites varying significantly between groups, i.e., the potential biomarkers of health condition, was carried out by search of the most intense MS1+ and MS1− signals of ROIMCR resolved components using HMDB database according to the criteria detailed in Section 2.6. Identification of metabolites, followed by identity confirmation through comparison of the assigned metabolite fragmentation pattern with the MS2 data simultaneously obtained by DDA, using the CFM-ID tool and MassBank database. This strategy allowed the identification of a total of 9, 41 and 36 compounds in the analyses A versus B, A versus C and B versus C groups, respectively. The complete list of identified metabolites (potential biomarkers), together with the associated ROIMCR component number, retention time (RT) in min, experimental m/z values, adduct ion, and associated error in ppm can be found in Tables S4–S6. The fold changes between the two analyzed groups, the type of regulation (up or down) and significance p values are also presented in Tables S4–S6. Two metabolites, at m/z 356.3833 (+) and m/z 633.7514 (+), could not be annotated in the searched databases. Interestingly, these two metabolites were only significantly different in A versus B and A versus C comparisons, being highly down-regulated in disease groups B and C (Tables S4 and S5). This result suggests that these features may be characteristic of control individuals (group A) and may be possible biomarkers of disease onset. Nevertheless, as their identification was not achieved yet, future studies should focus on confirming this result and on compound identification after, for instance, their previous purification and using other techniques. The analysis of Tables S4–S6 also evidenced that some of the identified metabolites could be associated with more than one ROIMCR component.

Different groups of compounds were found and identified, including amino acids and amino acid derivatives; different types of lipidic compounds such as lysophospholipids, fatty acid esters, and glycerophospholipids; bile acids; acyl carnitines; purine derivatives and uremic toxins (Tables S4–S6). Some of the compounds were observed to be up-regulated and others down-regulated in disease groups (B and C). The discovered metabolites and chemical groups are involved in several metabolic pathways that are known to be altered in chronic kidney disease (CKD), including amino acid metabolism, amine metabolism, lipid metabolism, steroid metabolism, the citric acid cycle, and the urea cycle [32,36]. A closer look on the identification of metabolites changing significantly between groups revealed molecules already recognized as biomarkers of CKD when group C was included in the comparison, such as the uremic toxins indoxyl sulfate, p-cresol sulfate, phenol sulfate, 4-hydroxybenzoic acid conjugates and hippuric acid, the dipeptide leucylproline, and creatinine, which is the classical marker of

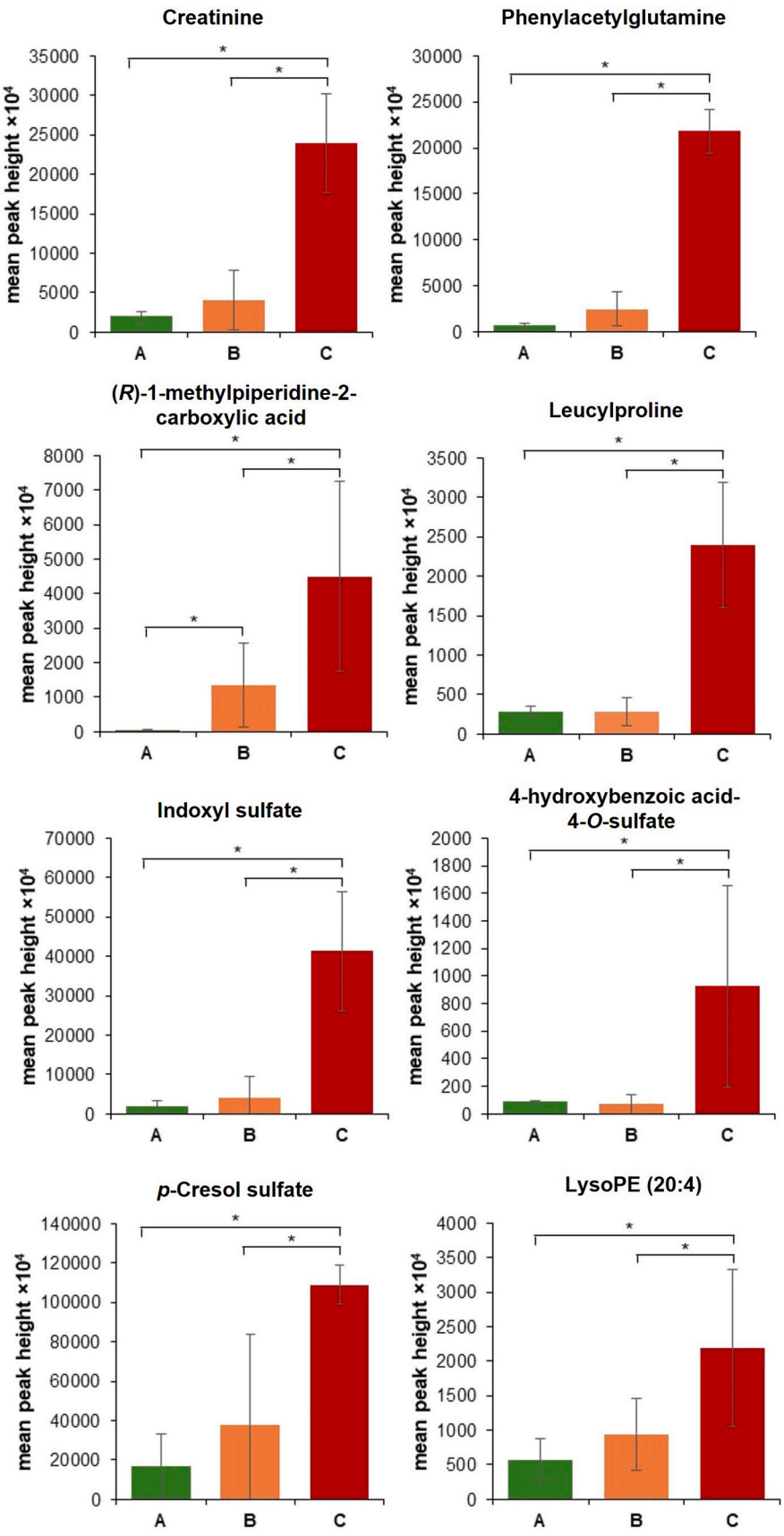

Fig. 5. Relative quantification of selected metabolites in the three groups under study: healthy controls (A), pre-dialysis CKD patients (B), and CKD patients at final stage (ESKD), undergoing peritoneal dialysis (C). Values are presented as mean peak height values of the elution profiles resolved by ROIMCR ± standard deviation. *, statistically significant differences observed in the comparison between groups ($p < 0.05$).

glomerular filtration rate (GFR) [32,33,36,52,54,55]. All these metabolites exhibited significantly higher levels in the final stage of disease (ESKD), when compared with healthy controls (analysis A versus C) or pre-dialysis patients, at an intermediate stage of disease (analysis B versus C) (Fig. 5). This result is in accordance with the accumulation of these compounds associated with the progressive decrease in renal function observed in CKD, and suggests the determination of their levels could be used to monitor disease progression [36,52,55]. The emerging gut microbiome-derived metabolite phenylacetyglutamine, which has been linked to a multitude of diseases involving renal, hepatic, cardiovascular, and cerebrovascular systems [56], was also significantly higher in end-stage patients, confirming its association with CKD (Fig. 5). Furthermore, the analysis of identified metabolites highlighted possible changes in the lipidic profile of plasma samples with the initiation and progression of CKD, being in line with previous studies describing a certain degree of dyslipidemia (Tables S4–S6, Fig. 5) [35,36].

Interestingly, the amino acid phenylalanine was only significantly lower in pre-dialysis disease patients (analysis A versus B) (Table S4), suggesting its possible application as early disease biomarker [33]. Moreover, the lysophosphatidylcholine LysoPC (20:5), the fatty acid 2,3-dimethylideneoctanedioic and the carboxylic acid derivative 2-methyl-2-phenoxypropanoic acid only varied significantly from the pre-dialysis (group B) to the final (group C) stage of disease (Table S6), indicating their possible use as disease progression biomarkers. Additionally, the alpha amino acid *(R)*-1-methylpiperidine-2-carboxylic acid varied significantly in all the performed comparisons (A versus B, A versus C, B versus C), with increasing levels observed in healthy controls (group A), pre-dialysis (group B) and end-stage (group C) CKD patients (Fig. 5). This result suggests that monitoring this metabolite could potentially be used as a biomarker of both the onset and the progression of disease. Nevertheless, the identification of metabolites performed in this work remains tentative (confidence level 2 according to the Metabolomics Standards Initiative [48] or confidence level 3 according to Schymanski et al. [49]) and should be further confirmed with authentic standards to fully verify the derived chemical structures. Moreover, targeted assays with relevant metabolites are recommended to accurately assess their levels in plasma samples.

Overall, the ROIMCR methodology applied to the analysis of plasma samples from healthy individuals and disease patients permitted to successfully identify metabolites already recognized as potential biomarkers of CKD, validating its application to pinpoint health-to-disease and disease progression indicators. Additionally, some metabolites varying significantly between health and disease groups that were not previously reported were now identified, evidencing the potential of ROIMCR to discover new biomarkers. The identification of compounds participating in lipid, carbohydrate, amino acid, and nucleic acid metabolism, pathways that are known to be altered in CKD, suggests that a panel of biomarkers, instead of single biomarkers, may constitute an improved strategy to accurately stage the disease and to better understanding the mechanisms underlying disease onset and progression. This is in line with recent metabolomic-based studies reporting the discovery of several metabolites highly correlated with mGFR and independent of demographic variables, which are being considered for a future panel of filtration markers to produce more accurate GFR estimations than the currently used creatinine [57,58]. Many of these metabolites were also identified in the present work including uremic toxins such as indoxyl sulfate and hippuric acid, myo-inositol lipids, alpha amino acids and amino acid derivatives like phenylacetylglutamine, reinforcing the usefulness of the herein proposed data processing strategy.

4. Conclusions

A methodology based in LC-HRMS and ROIMCR was successfully developed and applied to the untargeted metabolomic profiling of human plasma samples, in the search of health-to-disease biomarkers.

ROI procedure permitted the data compression and filtering, and the following MCR-ALS method resolved the elution profiles and spectra of the constituents present in plasma samples of healthy controls, intermediate stage CKD patients (pre-dialysis) and end-stage kidney disease patients (in dialysis). Positive and negative ionization data were simultaneously analyzed without the need of previous alignment of retention times between the two data sets which has never been reported before and reinforces the novelty of the present work. The integration of the MS1+ and MS1− spectra for each resolved component yielded an augmented metabolite coverage and facilitated their identification. Furthermore, the time required for data processing was reduced, resulting in a faster and more efficient analytical workflow.

Multivariate analysis evidenced that samples were separated and clustered according to their health status. Additionally, the metabolites responsible for groups differentiation have been identified, highlighting possible biomarkers. The number of metabolites varying significantly between groups was higher than described in previous studies on CKD research, particularly in the comparisons including end-stage patients, suggesting the potential of ROIMCR to find novel biomarkers and provide a more complete metabolomic profile, and representing an improvement compared to previous works. The identified metabolites changing between groups included compounds recognized as potential biomarkers of CKD, such as uremic toxins and creatinine, validating the application of ROICMR methodology for data processing of plasma samples in MS-based metabolomics studies. Furthermore, new potential biomarkers belonging to different chemical groups and involved in metabolic pathways that are known to be altered in CKD were discovered, confirming the potential of ROIMCR to find indicators of health condition.

Future studies will focus on confirming and investigating further these results by analysis of authentic standards of the identified metabolites and utilization of a higher number of samples per group. The application to alternative and non-invasive matrices, such as urine and saliva, and to different diseases, at distinct stages, is envisioned. Moreover, the metabolomic profiling of different biomatrices from the same individual along time will be also pursued as a body status fingerprint that can be linked to the pathogenesis of diseases, forming the basis of future personalized medicine strategies.

CRediT authorship contribution statement

Luisa Barreiros: Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Benedita Sampaio-Maia:** Writing – review & editing, Validation, Resources, Formal analysis. **Inês Soares Alencastre:** Writing – review & editing, Validation, Resources. **Romà Tauler:** Writing – review & editing, Supervision, Software, Resources, Methodology. **Marcela A. Segundo:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work received financial support from PT national funds (FCT/MECI, Fundação para a Ciência e a Tecnologia and Ministério da Educação, Ciência e Inovação) through project 2022.06012.PTDC DOI 10.54499/2022.06012.PTDC and project UID/50006 - Laboratório Associado para a Química Verde - Tecnologias e Processos Limpos. L. Barreiros acknowledges funding from FCT through program DL 57/2016 – Norma transitória. The HRMS analyses were performed using

equipment available at the Mass Spectrometry Laboratory of CEMUP, U. Porto, and the authors are grateful for the technical support provided. The authors thank the researchers of the Department of Environmental Chemistry at IDAEA-CSIC, particularly Flávia Yamamoto and Carlos Pérez-López, for their valuable help in successfully implementing the data processing methods. The authors also acknowledge Manuel Pestana (i3S, FMUP) for aiding in the biological interpretation of assigned metabolites.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.microc.2025.113737.

Data availability

Data will be made available on request.

References

[1] B. Amer, R.R. Deshpande, S.S. Bird, Simultaneous Quantitation and Discovery (SQUAD) Analysis: Combining the best of targeted and untargeted mass spectrometry-based metabolomics, Metabolites 13 (5) (2023) 648, https://doi.org/10.3390/metabo13050648.

[2] M. Roca, M.I. Alcoriza, J.C. Garcia-Canaveras, A. Lahoz, Reviewing the metabolome coverage provided by LC-MS: Focus on sample preparation and chromatography-A tutorial, Anal. Chim. Acta 1147 (2021) 38–55, https://doi.org/10.1016/j.aca.2020.12.025.

[3] T. Recber, E. Nemutlu, K. Beksac, O. Cennet, V. Kaynaroglu, S. Aksoy, S. Kir, Optimization and normalization strategies for long term untargeted HILIC-LC-qTOF-MS based metabolomics analysis: Early diagnosis of breast cancer, Microchem J. 179 (2022) 107658, https://doi.org/10.1016/j.microc.2022.107658.

[4] E. Gorrochategui, J. Jaumot, S. Lacorte, R. Tauler, Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow, TrAC-Trends Anal. Chem. 82 (2016) 425–442, https://doi.org/10.1016/j.trac.2016.07.004.

[5] J. Ivanisevic, E.J. Want, From samples to insights into metabolism: Uncovering biologically relevant information in LC-HRMS metabolomics data, Metabolites 9 (12) (2019) 308, https://doi.org/10.3390/metabo9120308.

[6] H.Y. Zhang, Z. Ouyang, W.P. Zhang, Advances in mass spectrometry for clinical analysis: Data acquisition, interpretation and information integration, TrAC-Trends Anal. Chem. 169 (2023) 117380, https://doi.org/10.1016/j.trac.2023.117380.

[7] S.M. Colby, C.H. Chang, J.L. Bade, J.R. Nunez, M.R. Blumer, D.J. Orton, K. J. Bloodsworth, E.S. Nakayasu, R.D. Smith, Y.M. Ibrahim, R.S. Renslow, T.O. Metz, DEIMoS: An open-source tool for processing high-dimensional mass spectrometry data, Anal. Chem. 94 (16) (2022) 6130–6138, https://doi.org/10.1021/acs.analchem.1c05017.

[8] Y. Ma, Y. Cao, X.C. Song, Y.Y. Zhang, J. Li, Y.K. Wang, X.Q. Wu, X.B. Qi, BAFinder: A software for unknown bile acid identification using accurate mass LC-MS/MS in positive and negative modes, Anal. Chem. 94 (16) (2022) 6242–6250, https://doi.org/10.1021/acs.analchem.1c05648.

[9] L. Naumann, A. Haun, A. Höchsmann, M. Mohr, M. Novák, D. Flottmann, C. Neusüss, Augmented region of interest for untargeted metabolomics mass spectrometry (AriumMS) of multi-platform-based CE-MS and LC-MS data, Anal. Bioanal. Chem. 415 (16) (2023) 3137–3154, https://doi.org/10.1007/s00216-023-04715-6.

[10] C. Pérez-López, A. Ginebreda, D. Barcelo, R. Tauler, SigSel: A MATLAB package for the pre and post-treatment of high-resolution mass spectrometry signals using the ROIMCR methodology, MethodsX 10 (2023) 102199, https://doi.org/10.1016/j.mex.2023.102199.

[11] C. Perez-Lopez, A. Ginebreda, J. Jaumot, F.Y. Yamamoto, D. Barcelo, R. Tauler, MSident: Straightforward identification of chemical compounds from MS-resolved spectra, Chemometr. Intell. Lab. Syst. 245 (2024) 105063, https://doi.org/10.1016/j.chemolab.2024.105063.

[12] C. Perez-Lopez, B. Oro-Noll, S. Lacorte, R. Tauler, Regions of interest multivariate curve resolution liquid chromatography with data-independent acquisition tandem mass spectrometry, Anal. Chem. 95 (19) (2023) 7519–7527, https://doi.org/10.1021/acs.analchem.2c05704.

[13] M.D.M. Delgado-Povedano, E. Maris, N. Kellner, G. Mulisa, L. Gámiz-Gracia, A. M. García-Campaña, M. De Boevre, S. De Saeger, R. Pero-Gascon, Liquid chromatography-tandem mass spectrometry for the determination of multiple mycotoxins in serum through suspect screening and targeted approaches: Advancing human mycotoxin biomonitoring, Microchem. J. 208 (2025) 112562. https://doi.org/10.1016/j.microc.2024.112562.

[14] A. Paul, P.D. Harrington, Chemometric applications in metabolomic studies using chromatography-mass spectrometry, TrAC-Trends Anal. Chem. 135 (2021) 116165, https://doi.org/10.1016/j.trac.2020.116165.

[15] E. Gorrochategui, J. Jaumot, R. Tauler, ROIMCR: a powerful analysis strategy for LC-MS metabolomic datasets, BMC Bioinf. 20 (2019) 256, https://doi.org/10.1186/s12859-019-2848-8.

[16] M. Perez-Cova, C. Bedia, D.R. Stoll, R. Tauler, J. Jaumot, MSroi: A pre-processing tool for mass spectrometry-based studies, Chemometr. Intell. Lab. Syst. 215 (2021) 104333, https://doi.org/10.1016/j.chemolab.2021.104333.

[17] M. Perez-Cova, J. Jaumot, R. Tauler, Untangling comprehensive two-dimensional liquid chromatography data sets using regions of interest and multivariate curve resolution approaches, TrAC-Trends Anal. Chem. 137 (2021) 116207, https://doi.org/10.1016/j.trac.2021.116207.

[18] A. Cerrato, C. Bedia, A.L. Capriotti, C. Cavaliere, V. Gentile, M. Maggi, C. M. Montone, S. Piovesana, A. Sciarra, R. Tauler, A. Lagan, Untargeted metabolomics of prostate cancer zwitterionic and positively charged compounds in urine, Anal. Chim. Acta 1158 (2021) 338381, https://doi.org/10.1016/j.aca.2021.338381.

[19] J.M. Jafari, J. Casas, C. Barata, H. Abdollahi, R. Tauler, Non-target ROIMCR LC-MS analysis of the disruptive effects of TBT over time on the lipidomics of Daphnia magna, Metabolomics 19 (8) (2023) 70, https://doi.org/10.1007/s11306-023-02030-w.

[20] M. Perez-Cova, S. Platikanov, R. Tauler, J. Jaumot, Quantification strategies for two-dimensional liquid chromatography datasets using regions of interest and multivariate curve resolution approaches, Talanta 247 (2022) 123586, https://doi.org/10.1016/j.talanta.2022.123586.

[21] C. Perez-Lopez, A. Ginebreda, M. Carrascal, D. Barcelo, J. Abian, R. Tauler, Non-target protein analysis of samples from wastewater treatment plants using the regions of interest-multivariate curve resolution (ROIMCR) chemometrics method, J. Environ. Chem. Eng. 9 (4) (2021) 105752, https://doi.org/10.1016/j.jece.2021.105752.

[22] C. Perez-Lopez, S. Rodriguez-Mozaz, A. Serra-Compte, D. Alvarez-Munoz, A. Ginebreda, D. Barcelo, R. Tauler, Effects of sulfamethoxazole exposure on mussels (Mytilus galloprovincialis) metabolome using retrospective non-target high-resolution mass spectrometry and chemometric tools, Talanta 252 (2023) 123804, https://doi.org/10.1016/j.talanta.2022.123804.

[23] F.Y. Yamamoto, C. Perez-Lopez, A. Lopez-Antia, S. Lacorte, D.M.D. Abessa, R. Tauler, Linking MS1 and MS2 signals in positive and negative modes of LC-HRMS in untargeted metabolomics using the ROIMCR approach, Anal. Bioanal. Chem. 415 (2023) 6213–6225, https://doi.org/10.1007/s00216-023-04893-3.

[24] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and applications, Chemometr. Intell. Lab. Syst. 140 (2015) 1–12, https://doi.org/10.1016/j.chemolab.2014.10.003.

[25] R. Tauler, Multivariate curve resolution applied to second order data, Chemometr. Intell. Lab. Syst. 30 (1) (1995) 133–146, https://doi.org/10.1016/0169-7439(95)00047-x.

[26] L.L. Hohrenk, M. Vosough, T.C. Schmidt, Implementation of chemometric tools to improve data mining and prioritization in LC-HRMS for nontarget screening of organic micropollutants in complex water matrixes, Anal. Chem. 91 (14) (2019) 9213–9220, https://doi.org/10.1021/acs.analchem.9b01984.

[27] R. Tauler, Multivariate curve resolution of multiway data using the multilinearity constraint, J. Chemometr. 35 (2) (2021) e3279, https://doi.org/10.1002/cem.3279.

[28] A. Queral-Beltran, M. Marin-Garcia, S. Lacorte, R. Tauler, UV-Vis absorption spectrophotometry and LC-DAD-MS-ESI(+)-ESI(-) coupled to chemometrics analysis of the monitoring of sulfamethoxazole degradation by chlorination, photodegradation, and chlorination/ photodegradation, Anal. Chim. Acta 1276 (2023) 341563, https://doi.org/10.1016/j.aca.2023.341563.

[29] M. Perez-Cova, R. Tauler, J. Jaumot, Adverse effects of arsenic uptake in rice metabolome and lipidome revealed by untargeted liquid chromatography coupled to mass spectrometry (LC-MS) and regions of interest multivariate curve resolution, Separations 9 (3) (2022) 79, https://doi.org/10.3390/separations9030079.

[30] J.C. Lv, L.X. Zhang, Prevalence and disease burden of chronic kidney disease, in: B. C. Liu, H.Y. Lan, L.L. Lv (Eds.), Renal Fibrosis: Mechanisms and Therapies, Springer, Singapore, 2019, pp. 3–15.

[31] T.H. Yeh, K.C. Tu, H.Y. Wang, J.Y. Chen, From acute to chronic: unraveling the pathophysiological mechanisms of the progression from acute kidney injury to acute kidney disease to chronic kidney disease, Int. J. Mol. Sci. 25 (3) (2024) 1755, https://doi.org/10.3390/ijms25031755.

[32] X. Fu, Z.X. Luo, H.H. Yin, Y.N. Liu, X.G. Du, W. Cheng, J.Y. Liu, Metabolomics study reveals blood biomarkers for early diagnosis of chronic kidney disease and IgA nephropathy: A retrospective cross-sectional study, Clin. Chim. Acta 555 (2024) 117815, https://doi.org/10.1016/j.cca.2024.117815.

[33] M.R. Glavan, C. Socaciu, A.I. Socaciu, F. Gadalean, O.M. Cretu, A. Vlad, D. M. Muntean, F. Bob, O. Milas, A. Suteanu, D.C. Jianu, M. Stefan, L. Balint, S. Ienciu, L. Petrica, Untargeted metabolomics by ultra-high-performance liquid chromatography coupled with electrospray ionization-quadrupole-time of flight-mass spectrometry analysis identifies a specific metabolomic profile in patients with early chronic kidney disease, Biomedicines 11 (4) (2023) 1057, https://doi.org/10.3390/biomedicines11041057.

[34] M. Kordalewska, S. Macioszek, R. Wawrzyniak, M. Sikorska-Wisniewska, T. Sledzinski, M. Chmielewski, A. Mika, M.J. Markuszewski, Multiplatform metabolomics provides insight into the molecular basis of chronic kidney disease, J. Chromatogr. B 1117 (2019) 49–57, https://doi.org/10.1016/j.jchromb.2019.04.003.

[35] N. Ragi, K. Sharma, Deliverables from metabolomics in kidney disease: adenine, new insights, and implication for clinical decision-making, Am. J. Nephrol. 55 (4) (2024) 421–438, https://doi.org/10.1159/000538051.

[36] Y.N. Wang, S.X. Ma, Y.Y. Chen, L. Chen, B.L. Liu, Q.Q. Liu, Y.Y. Zhao, Chronic kidney disease: Biomarker diagnosis to therapeutic targets, Clin. Chim. Acta 499 (2019) 54–63, https://doi.org/10.1016/j.cca.2019.08.030.
[37] D. Broadhurst, R. Goodacre, S.N. Reinke, J. Kuligowski, I.D. Wilson, M.R. Lewis, W. B. Dunn, Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies, Metabolomics 14 (6) (2018) 72, https://doi.org/10.1007/s11306-018-1367-3.
[38] M.C. Chambers, B. Maclean, R. Burke, D. Amodei, D.L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T.A. Baker, M.Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S.L. Seymour, L.M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E.W. Deutsch, R. L. Moritz, J.E. Katz, D.B. Agus, M. MacCoss, D.L. Tabb, P. Mallick, A cross-platform toolkit for mass spectrometry and proteomics, Nat. Biotechnol. 30 (10) (2012) 918–920, https://doi.org/10.1038/nbt.2377.
[39] L.B. Cao, P.D. Harrington, J.D. Liu, SIMPLISMA and ALS applied to two-way nonlinear wavelet compressed ion mobility spectra of chemical warfare agent simulants, Anal. Chem. 77 (8) (2005) 2575–2586, https://doi.org/10.1021/ac0486286.
[40] R. Bro, A.K. Smilde, Principal component analysis, Anal. Methods 6 (9) (2014) 2812–2831, https://doi.org/10.1039/c3ay41907j.
[41] M. Barker, W. Rayens, Partial least squares for discrimination, J. Chemometr. 17 (3) (2003) 166–173, https://doi.org/10.1002/cem.785.
[42] A.K. Smilde, J.J. Jansen, H.C.J. Hoefsloot, R. Lamers, J. van der Greef, M. E. Timmerman, ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data, Bioinformatics 21 (13) (2005) 3043–3048, https://doi.org/10.1093/bioinformatics/bti476.
[43] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemometr. Intell. Lab. Syst. 58 (2) (2001) 109–130, https://doi.org/10.1016/s0169-7439(01)00155-1.
[44] D.S. Wishart, A.C. Guo, E. Oler, F. Wang, A. Anjum, H. Peters, R. Dizon, Z. Sayeeda, S.Y. Tian, B.L. Lee, M. Berjanskii, R. Mah, M. Yamamoto, J. Jovel, C. Torres-Calzada, M. Hiebert-Giesbrecht, V.W. Lui, D. Varshavi, D. Varshavi, D. Allen, D. Arndt, N. Khetarpal, A. Sivakumaran, K. Harford, S. Sanford, K. Yee, X. Cao, Z. Budinski, J. Liigand, L. Zhang, J.M. Zheng, R. Mandal, N. Karu, M. Dambrova, H.B. Schiöth, R. Greiner, V. Gautam, HMDB 5.0: the human metabolome database for 2022, Nucl. Acids Res. 50(D1) (2022) D622–D631, https://doi.org/10.1093/nar/gkab1062.
[45] D.S. Wishart, D. Tzur, C. Knox, R. Eisner, A.C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M.A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D.D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G. E. Duggan, G.D. MacInnis, A.M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B.D. Sykes, H.J. Vogel, L. Querengesser, HMDB: the human metabolome database, Nucl. Acids Res. 35 (2007) D521–D526, https://doi.org/10.1093/nar/gkl923.
[46] F. Wang, D. Allen, S.Y. Tian, E. Oler, V. Gautam, R. Greiner, T.O. Metz, D.S. Wishart, CFM-ID 4.0-a web server for accurate MS-based metabolite identification, Nucl. Acids Res. 50(W1) (2022) W165–W174, https://doi.org/10.1093/nar/gkac383.
[47] F. Wang, J. Liigand, S.Y. Tian, D. Arndt, R. Greiner, D.S. Wishart, CFM-ID 4.0: More accurate ESI-MS/MS spectral prediction and compound identification, Anal. Chem. 93(34) (2021) 11692-11700, https://doi.org/10.1021/acs.analchem.1c01465.
[48] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W. M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane, J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reily, J.J. Thaden, M.R. Viant, Proposed minimum reporting standards for chemical analysis, Metabolomics 3 (3) (2007) 211–221, https://doi.org/10.1007/s11306-007-0082-2.
[49] E.L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H.P. Singer, J. Hollender, Identifying small molecules via high resolution mass spectrometry: Communicating confidence, Environ. Sci. Technol. 48 (4) (2014) 2097–2098, https://doi.org/10.1021/es5002105.
[50] A.M. Knolhoff, C.N. Kneapler, T.R. Croley, Optimized chemical coverage and data quality for non-targeted screening applications using liquid chromatography/high-resolution mass spectrometry, Anal. Chim. Acta 1066 (2019) 93–101. https://doi.org/10.1016/j.aca.2019.03.032.
[51] S. Benito, N. Unceta, M. Maciejczyk, A. Sánchez-Ortega, K. Taranta-Janusz, J. Szulimowska, A. Zalewska, F. Andrade, A. Gómez-Caballero, P. Dubiela, R. J. Barrio, Revealing novel biomarkers for diagnosing chronic kidney disease in pediatric patients, Sci. Rep. 14 (1) (2024) 11549, https://doi.org/10.1038/s41598-024-62518-w.
[52] H. Lee, K.H. Liu, Y.H. Yang, J.D. Liao, B.S. Lin, Z.Z. Wu, A.C. Chang, C.C. Tseng, M. C. Wang, Y.S. Tsai, Advances in uremic toxin detection and monitoring in the management of chronic kidney disease progression to end-stage renal disease, Analyst 149 (10) (2024) 2784–2795, https://doi.org/10.1039/d4an00057a.
[53] R. Dator, P.W. Villalta, N. Thomson, J. Jensen, D.K. Hatsukami, I. Stepanov, B. Warth, S. Balbo, Metabolomics profiles of smokers from two ethnic groups with differing lung cancer risk, Chem. Res. Toxicol. 33 (8) (2020) 2087–2098, https://doi.org/10.1021/acs.chemrestox.0c00064.
[54] Y. Chen, L.R. Zelnick, K. Wang, A.N. Hoofnagle, J.O. Becker, C.Y. Hsu, H. I. Feldman, R.C. Mehta, J.P. Lash, S.S. Waikar, T. Shafi, S.L. Seliger, M.G. Shlipak, M. Rahman, B.R. Kestenbaum, C.S. Investigators, Kidney clearance of secretory solutes is associated with progression of CKD: The CRIC study, J. Am. Soc. Nephrol. 31 (4) (2020) 817–827, https://doi.org/10.1681/asn.2019080811.
[55] Y.R. Ma, M.Y. Xin, K. Li, H. Wang, Z. Rao, T.X. Liu, X.A. Wu, An LC-MS/MS analytical method for the determination of uremic toxins in patients with end-stage renal disease, J. Pharm. Biomed. Anal. 191 (2020) 113551, https://doi.org/10.1016/j.jpba.2020.113551.
[56] N.K. Krishnamoorthy, M. Kalyan, T.A. Hediyal, N. Anand, P.H. Kendaganna, G. Pendyala, S.V. Yelamanchili, J. Yang, S.B. Chidambaram, M.K. Sakharkar, A. M. Mahalakshmi, Role of the gut bacteria-derived metabolite phenylacetylglutamine in health and diseases, ACS Omega 9 (3) (2024) 3164–3172, https://doi.org/10.1021/acsomega.3c08184.
[57] N.F. Fino, O.M. Adingwupu, J. Coresh, T. Greene, B. Haaland, M.G. Shlipak, V. T. Silva, R. Kalil, A.L. Mindikoglu, S.L. Furth, J.C. Seegmiller, A.S. Levey, L. A. Inker, Evaluation of novel candidate filtration markers from a global metabolomic discovery for glomerular filtration rate estimation, Kidney Int. 105 (3) (2024) 582–592, https://doi.org/10.1016/j.kint.2023.11.007.
[58] T.B. Li, M.E. Grams, L.A. Inker, J.S. Chen, E.P. Rhee, B.A. Warady, A.S. Levey, M. R. Denburg, S.L. Furth, V.S. Ramachandran, P.L. Kimmel, J. Coresh, Consistency of metabolite associations with measured glomerular filtration rate in children and adults, Clin. Kidney J. 17 (6) (2024) sfae108, https://doi.org/10.1093/ckj/sfae108.