



A Nontargeted UHPLC-HRMS Metabolomics Pipeline for Metabolite Identification: Application to Cardiac Remote Ischemic Preconditioning

Judith Kouassi Nzoughet, Cinzia Bocca, Gilles Simard, Delphine Prunier-Mirebeau, Juan Manuel Chao de la Barca, Dominique Bonneau, Vincent Procaccio, Fabrice Prunier, Guy Lenaers, Pascal Reynier

► To cite this version:

Judith Kouassi Nzoughet, Cinzia Bocca, Gilles Simard, Delphine Prunier-Mirebeau, Juan Manuel Chao de la Barca, et al.. A Nontargeted UHPLC-HRMS Metabolomics Pipeline for Metabolite Identification: Application to Cardiac Remote Ischemic Preconditioning. *Analytical Chemistry*, American Chemical Society, 2017, 89 (3), pp.2138-2146. 10.1021/acs.analchem.6b04912 . hal-02544712

HAL Id: hal-02544712

<https://hal.archives-ouvertes.fr/hal-02544712>

Submitted on 12 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution| 4.0 International License

A non-targeted UHPLC-HRMS metabolomics pipeline for metabolite identification; application to cardiac remote ischemic preconditioning

Judith Kouassi Nzoughe^{1*}, Cinzia Bocca¹, Gilles Simard^{1,2}, Delphine Prunier-Mirebeau^{1,2}, Juan Manuel Chao de la Barca^{1,2}, Dominique Bonneau^{1,2}, Vincent Procaccio^{1,2}, Fabrice Prunier^{3,4}, Guy Lenaers¹, Pascal Reynier^{1,2}.

¹PREMMi, Pôle de Recherche et d'Enseignement en Médecine Mitochondriale, Institut MITOVASC, CNRS 6214, INSERM U1083, Université d'Angers, 4 rue Larrey, 49933 Angers cedex 9, France ;

²Département de Biochimie et Génétique, Centre Hospitalier Universitaire, Angers, France ;

³Institut MITOVASC, Laboratoire EA3860, Cardioprotection, remodelage et thrombose, Angers, France ;

⁴Département de Cardiologie, Centre Hospitalier Universitaire, Angers, France.

Abstract

In recent years, the amount of investigations based on non-targeted metabolomics has increased, although often without thorough assessment of analytical strategies applied to acquire data. Following published guidelines for metabolomics experiments, we report a validated non-targeted metabolomics strategy with pipeline for unequivocal metabolites identification using the MSMLS™ molecule library. We achieved an in-house database containing accurate m/z values, retention times, isotopic patterns, full MS and MS/MS spectra. A UHPLC-HRMS Q-Exactive™ method was developed and experimental variations were determined within and between 3 experimental days. The extraction efficiency as well as the accuracy, precision, repeatability, and linearity of the method were assessed, the method demonstrating good performances. The methodology was further blindly applied to plasma from Remote Ischemic Pre-Conditioning (RIPC) rats. Samples, previously analyzed by targeted metabolomics using completely different protocol, analytical strategy and platform, were submitted to our analytical pipeline. A combination of multivariate and univariate statistical analyses was employed. Selection of putative biomarkers from OPLS-DA model and S-plot was combined to jack-knife confidence intervals, metabolites VIP values and univariate statistics. Only variables with

strong model contribution and highly statistical reliability were selected as discriminated metabolites. Three biomarkers identified by the previous targeted metabolomics study were found in the current work, in addition to three novel metabolites, emphasizing the efficiency of the current methodology and its ability to identify new biomarkers of clinical interest, in a single sequence. The biomarkers were identified to level 1 according to the Metabolomics Standard Initiative and confirmed by both RPLC and HILIC-HRMS.

INTRODUCTION

Metabolomics is defined as the comprehensive analysis of low-molecular-weight metabolites, typically <1500 Daltons, which are highly context dependent, varying according to the physiology, developmental or pathological state of a cell, tissue, organ or organism ¹. Two major analytical techniques are mostly used: nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS), with the latter becoming more widely exploited in the field ². In biomarker discovery, MS has the advantage of very high sensitivity and the ability to detect a high number of different metabolites depending on the experimental setup. Moreover, the specificity of MS, through high resolution and/or multidimensional MSⁿ techniques, further facilitates the structural elucidation of metabolites of interest ^{3,4}. In the recent years, the amount of clinical investigations based on metabolomics has considerably increased, although often without thorough assessment of the analytical methods applied to acquire data especially in non-targeted metabolomics.

Metabolomics needs highly standardized methods to avoid bias and data misinterpretations. Thus, efforts have been made for defining appropriate validation parameters. Challenges in non-targeted strategies are different to targeted methods for which published guidelines exist ⁵⁻⁸. Targeted metabolomics approaches focus on the quantification of a limited number of well characterized pre-selected molecules. In contrast, non-targeted metabolomics methods are more exhaustive but they often lack the undeniable characterization of the metabolites of interest. In 2007, minimum meta data relative to instrumental performance and method validation have been proposed by the Metabolomics Standards Initiative (MSI) ⁹, followed by minimum reporting standards for data analysis associated with metabolomics experiments ¹⁰. More recently, a relevant review compiling alternative approaches used to validate metabolomics methodologies was published ¹¹ and further validation criteria for non-targeted metabolomics were recommended. Criteria include the assessment of: the accuracy and precision for selected compounds with different physico-chemical properties, retention time, pooled QCs to measure the repeatability within an analytical run and filtering data before analysis (considering the drift in signal variations), diluted pooled QCs for checking linearity, and total signal plot from each chromatogram to verify instrumental repeatability.

With the aim of setting up clinical metabolic profiling strategy, the above-mentioned recommendations were considered in order to provide high degree of confidence in the developed

method and to ensure that the methodologies were fit for purpose. A method based on Ultra High Performance Liquid Chromatography coupled to High Resolution Mass Spectrometry (UHPLC-HRMS) Q-Exactive™ was developed and subjected to a validation process. To overcome the difficulties of metabolites identification and unequivocally biomarkers identification, an extended library of more than 500 molecules intended for mass spectrometry metabolomics applications was used. Identification and confirmation of biomarkers from metabolomics investigations are essential. Further, the methodology was blindly applied to plasma from RIPC rats, previously analyzed by completely different targeted metabolomics protocol, analytical strategy and platform¹².

The use of the same sample cohort in the current research had several purposes: (1) to validate the developed workflow, (2) to test the in-house database and (3) to see whether similar biomarkers can be obtained using different analytical platforms and technologies.

EXPERIMENTAL SECTION

A schematic diagram of the experimental design is illustrated in Figure 1.

Chemicals and reagents

Description of the solvents, chemicals, and authentic standards used can be found in Supporting Information.

Study design and animal experiment

This is a 2-steps study, with the first part related to the validation of the developed method, the creation of the in-house molecules library, and a second part dedicated to the implementation of the methodology and its application to a RIPC clinical cohort.

Human plasma samples were obtained from Angers University Hospital. These plasmas were collected from consented patients attending clinics. One mL plasma from four patients was pooled together and constituted sample matrix for the method development and validation. Besides, plasma from rats involved in a controlled RIPC experimentation was further used to assess the potential of the developed strategies for clinical metabolomics and biomarkers identification.

The RIPC animal experimentation has been previously described¹³. Briefly, 20 male Wistar rats, 8- to 10-weeks old, were randomly assigned to either RIPC group (T) or control group (C). Plasma was collected using standard procedures and stored at -80°C prior to use. Further details on the RIPC procedure can be found in Supporting Information.

Plasma sample extraction

Due to the wide chemical diversity of metabolites, a complete holistic overview using metabolomics is not possible. The sample extraction procedure employed herein uses minimal sample treatment and was expected to be non-selective in order to reveal as much information as possible. Plasmas were

extracted with ice-cold MeOH using standard methods for non-targeted metabolomics using LC-HRMS. A detailed extraction protocol is available in Supporting Information.

For the RPC investigation, a pooled quality control sample deriving from all rat subjects was prepared, to ensure that no or minimal metabolic information was lost¹⁴. QC dilution series (1:2, 1:4 in reconstitution solvent) were also carried out and provided robust quality assurance for each metabolic feature detected.

UHPLC RP- and HILIC-HRMS Q-Exactive™ analysis

A Dionex™ UltiMate™ 3000 UHPLC system (Dionex, Sunnyvale, USA) coupled to a Thermo Scientific Q Exactive™ HRMS (Thermo Fisher Scientific, Bremen, Germany) equipped with a heated electrospray (HESI II) source was used for metabolites analysis and accurate mass measurements.

Reverse phase (RP) conditions combined to electrospray ionization were applied, as recognized to be more open in terms of application range. The chromatographic separation was achieved with an Acquity® CSH C18 1,7µm 150 mm × 2,1 column together with the corresponding pre-column CSH C18 1,7µm VanGuard (Waters, Guyancourt, France). The analytical column and the autosampler were maintained at a temperature of 40°C and 4°C, respectively. Mobile phases consisted of H₂O in channel A and MeOH in channel B, both containing 0.1 % formic acid. The elution gradient (A : B, v/v) was as follow: hold initial conditions 98:2 for 2 min, followed by a linear gradient from 98:2 to 0:100 over a 15 min period, hold at 0:100 for 3 min, return to initial conditions 98:2 over 2.5 min and then hold these conditions for a further 2 min. A constant flow rate of 0.300 mL min⁻¹ was used; the injection volume was optimized at 5 µl and samples injection order was randomized. A divert valve was used and the eluent was directed to waste at 22.45 min. Between each injection, the system was equilibrated for 1.5 min.

High resolution MS was acquired in positive and negative ionization mode, respectively (a distinct run for each modality). Full scan mass spectra (Full MS) were acquired and data-dependent MS/MS (ddMS2) experiments were performed at the start and the end of each sequence, on several QCs, and acquired in 'Top5' data-dependent mode. MS conditions can be found in Supporting Information. Xcalibur 2.2 software (Thermo Fisher Scientific, San Jose, CA, USA) was used for data acquisition. Prior to each sequence acquisition, the mobile phase was run for 1h30, followed by injection of 3 solvent blanks and 5 QC samples to allow column equilibration and conditioning. Column pressures and isotope labeled standards monitoring were performed at every analytical run for verifying the current state of the whole system, as a quality control procedure.

A Hydrophilic Interaction Chromatography (HILIC) –HRMS method was further implemented to ascertain the biomarkers identified by RPLC-HRMS. The chromatographic separation was achieved with an Acquity® BEH HILIC 1.7µm 150x2.1mm column together with the corresponding pre-

column BEH HILIC 1,7 μ m VanGuard (Waters, Guyancourt, France). The analytical column and the autosampler were maintained at 35°C and 4°C, respectively. Mobile phases consisted of H₂O in channel A and ACN/H₂O (99/1) in channel B, both containing 10mM ammonium acetate and 0.05 % acetic acid. The system was programmed to perform an analysis cycle consisting of holding initial conditions 95% B for 2 min, followed by a linear gradient from 95% to 80% B over a 3 min period, from 80% to 60% B over 7 min, from 60% to 40% B over 2 min, hold at 40% B for 2 min, return to initial conditions over 2 min and then hold of these conditions for a further 10 min. The flow rate was 0.400 mL min⁻¹ and the injection volume was 10 μ l. High resolution MS was acquired in positive ionization mode. Targeted SIM (t-SIM) and targeted MS2 (t-MS2) experiments were performed for each marker of interest. MS conditions are detailed in Supporting Information.

Validation procedure

There are no guidelines for validating analytical methods in non-targeted approach. Nonetheless, the following articles constitute adequate references for conducting method validation. Indeed, analytical methods used for non-targeted metabolomics investigations should be thoroughly assessed prior to their use. Validation of the analytical method took into account the suggestions of the MSI⁹, recent propositions for method validation in non-targeted metabolomics¹¹ and considered also conventional guidelines for quantitative methods. The method was validated in terms of extraction recovery, selectivity, repeatability, method precision, linearity, and instrumental precision. Limit of detection (LOD) and limit of quantification (LOQ) of the internal standards were also assessed, based on signal to noise ratios at 3 and at 10, respectively.

Extraction recovery

Pooled human plasmas were used to estimate the extraction recovery; to this end, samples were spiked at start (six independent replicates, QC_SS) and at the end of the extraction procedure (six independent replicates, QC_SE) with isotope labeled compounds exhibiting different functional groups, polarities, and molecular masses. Fortified materials were extracted following the procedure described above. This procedure was performed over three different days and for each ionization mode (positive and negative, respectively). The extraction recovery was determined by comparing the peak area of isotope labeled compounds in pre-fortified test material extract (QC_SS, fortified before extraction procedure) with post-fortified extract (QC_SE, fortified after extraction).

Selectivity

Selectivity of a method refers to the extent to which it can determine particular analyte(s) in a complex mixture without interference from other components in the mixture⁽¹⁵⁾. Here, selectivity was provided by detection of metabolites in the matrix, based on compounds' exact masses.

Repeatability and method precision (within-day, between-day and intermediate precision)

One-way analysis of variance (ANOVA) was applied, not to test significance between-day difference, but to correctly estimate variances and generate assay precision coefficients of variation (CV). The within-day precision, between-day and intermediate precision were calculated as computed in equation 1, equation 2 and equation 3, in Supporting Information.

Linearity

Serial dilutions of QC extracts (1:2 and 1:4) were used to assess the linearity of the response for isotope labeled standards.

Instrumental precision

Instrumental precision was assessed by evaluating analytical reproducibility (intensity and peak area accuracy), system stability (mass accuracy, chromatograms alignment) and chromatographic reproducibility (retention time accuracy, chromatograms alignment). Chromatograms intensities were further plotted against all samples involved through the validation process, i.e., at day1, day2, and day3 for each ionization mode.

Data processing

SIEVE™ v2.2 software (Thermo Fisher Scientific) was used to preprocess ‘.raw’ files from the UHPLC-HRMS. The algorithm ‘Component’ was applied for background subtraction, component detection, peak alignments and framing. This step was meant to ensure the validity of the sequence, with appropriate chromatogram alignment before further processing. TraceFinder™3.0 software (Thermo Fisher Scientific) was employed for data processing, as it allowed us to verify automatically: isotopic pattern, expected m/z vs. experimental m/z, expected retention time (RT) vs. experimental RT, peak integration, dilutions linearity (1, 1/2, 1/4 dilutions), MS/MS fragments (Top5 ddMS2 experiments were initiated for RIPC rat cohort), comparison with the MSMLS™ library for database match and metabolites identification. Inter-batch normalization was applied to data from day1, day2 and day3 of the validation process; each sampling day corresponding to one analytical batch (sequence). The normalization followed the procedure previously described, assuming that the measurement errors in a single batch are randomly distributed and that different batches can be compared and corrected using the average or median value of the QC samples in a batch ¹⁶.

Metabolites identification

MSMLS™ is a collection of high quality small biochemical molecules that span a broad range of primary metabolism. From the MSMLS™ metabolites library, an in-house database (internal

metabolites library) was created. Mass spectra were compared to those available in mzcloud and HMDB¹ mass spectral databases. Biomarkers identification was therefore facilitated using our in-house library of 500 accurately identified and well characterized molecules. Accurate m/z, Full MS, MS/MS spectra, RT and isotopic pattern of metabolites were acquired on the same platform using the current chromatographic and mass spectrometry conditions. The metabolite library used in this study therefore contains accurate m/z values, retention times, Isotopic pattern, Full MS and MS/MS spectra for each metabolites, to facilitate identification.

A processing method was created using TraceFinder™ 3.0 software. The generated data matrix containing identified features was filtered out based on the following criteria: CV below 30%, accurate m/z measurement with Δ ppm < 5, isotopic pattern (masses and abundances of the isotopes, pass), expected retention time (RT, pass), linearity for dilutions 1, 1/2 and 1/4, library search (pass) and MS/MS fragments (Top5 ddMS2 experiments) database matching. Estimation of RT drift was made from RT time of the metabolite in the internal library vs. experimental RT of metabolite during the sequence; Δ RT \pm 10s was used. Metabolites identification required a manual checking of each feature in TraceFinder™, isomers selection by visualization of RT and MS2 fragmentations before exporting the generated data matrix for multivariate and univariate statistics.

Data Analysis

Data were exported to Microsoft Excel® for inter-batch normalization, then to Simca P+ v 14.0 software (Umetrics, Umea, Sweden) for multivariate statistical analysis, and finally to SPSS Statistics v22 (IBM, Bois-Colombes, France) for univariate statistics.

Multivariate statistical analysis

Data were subjected to Log transformation and Pareto scaling prior to Principal Component Analysis (PCA) and Orthogonal Projection of Latent Structures-Discriminant Analysis (OPLS-DA). This was followed by an S-plot which provided examination of the OPLS-DA predictive component loading to facilitate model interpretation. Selection of putative biomarkers from the S-plot was combined to jack-knife confidence intervals from a loading column plot, and looking at the Variable Importance in the Projection (VIP) values of those variables. Consequently, only metabolites with strong model contribution and highly statistical reliability were selected as discriminated metabolites (putative biomarkers).

The quality of the finally obtained OPLS-DA model was evaluated by R² (goodness of fit, i.e., how well the model fits the data), Q² (goodness of prediction, i.e., how well the model predicts new data) parameters, cross validation-analysis of variance CV-ANOVA¹⁷ and a permutation test. CV-ANOVA

is a diagnostic tool for assessing the reliability of OPLS models; the returned p-value is indicative of the statistical significance of the investigated model.

Univariate statistical analysis

RT and intensity CVs were computed in Microsoft Excel[®] for each labeled isotope standard as for the ratio of abundance observed between C/T and T/C samples for each metabolite. ANOVA test was applied to the validation dataset in order to generate repeatability precision coefficients of variation.

CV within QC samples and p-value further served for metabolites selection in the RPC cohort.

Kolmogorov-Smirnov and Shapiro-Wilk tests were performed to test for normality, whilst Levene's test was used to assess the homogeneity of variance. Unequal variance two-tailed t-test was then performed to assess the significance of biomarkers of interest in this experiment. The significance threshold was set at $p < 0.05$. Metabolites of interest were confirmed by MS/MS spectral database, RT and isotopic pattern matching. This workflow was meant to allow a consistent exploration of metabolomics signatures with identification and confirmation of associated biomarkers in a single experimental sequence.

Safety considerations

Experiments were carried out according to all relevant ethics and institutional guidelines. COSHH/Risk Assessment Forms were completed before carrying out any research experiments. Appropriate personal protection equipment was used and chemicals were handled in a fume hood. Flammable items were kept in chemical safety cabinet and material wastes were disposed of through clinical wastes.

RESULTS AND DISCUSSION

Method Performances

As quality control procedure, the following practices were systematically adopted before any data processing and analysis: comparison of column pressure to the previous analytical run, internal standards variation in QCs/study samples, and instrumental stability (reproducibility of RT and accurate masses) check along the sequence. These observations served as grounds for the acceptance of the analytical run. This is further reinforced through peaks alignment (Figure S-1) confirming chromatographic reproducibility and system stability of the whole system.

The isotope labeled endogenous metabolites used were of diverse nature and chemical structure, covering a broad range of molecular masses, functional groups, polarities, and with RT covering the

entire chromatogram. These include: 17 α -Hydroxyprogesterone-D8 (m/z: 339.27699; RT: 16.97 min), DL-Alanine-15N (m/z: 91.05199; RT: 1.20 min), L-Thyroxine-13C6 (m/z: 783.71409; RT: 16.22 min) in positive mode; Succinic acid-2,2,3,3-d4 (m/z: 121.04334; RT: 2.54 min), Pyruvic acid-1-13C (m/z: 88.01103; RT: 3.36 min), and L-Thyroxine-13C6 (m/z: 781.69844; RT: 16.22 min) in negative mode.

The objective of spiking with isotope labelled endogenous metabolites was to obtain a general snapshot of the method, even though results cannot be formally extended to all metabolites in the sample. The analytical performances of the method are presented in Table S-1. As can be depicted from Table S-1, the method displays good extraction recoveries (comprised between 70 and 84%), within-day and between-day CV below 10%, and excellent instrumental precisions. The acceptance criteria for method repeatability were set at 15%, according to the FDA recommendations⁷.

Peak area and intensity repeatability for isotope labeled standards was set at 30%; it is common in non-targeted metabolomics analysis by LC-MS to proceed with ions exhibiting CV below 30% in QCs, since it is considered that ions with higher CV would not be good candidates as biomarkers. Checking the linearity of metabolites in diluted QCs was evaluated during the validation study, as we were planning to use the criteria of linear trend to filter metabolites in the data matrix. We only performed 1/2 and 1/4 dilutions as we considered that further dilutions might lead to exclusion of some potential metabolites of low abundance.

In-house database for metabolites identification

The MSMLS™ library contains 619 unique metabolites, of a broad spectrum of key primary metabolites and intermediates including the following classes of compounds: carboxylic acids, amino acids, biogenic amines, polyamines, nucleotides, coenzymes, vitamins, mono - and disaccharides, fatty acids, lipids, steroids, and hormones. Among the 619 metabolites of MSMLS™ library, we were able to reliably analyze 499 metabolites, in positive and negative ionization mode (Table S-2). Urea, initially not contained in the MSMLS™ library was added to our database. Our in-house database includes 500 metabolites of key pathways, for which full MS, MS/MS spectra, RT and isotopic pattern were acquired, under our current chromatographic and mass spectrometry conditions.

Identification and confirmation of biomarkers from metabolomics investigations are essential for precision medicine.

In the case of isomer metabolites emerging as putative biomarkers, and not chromatographically separated, only the chemical formula is given. A specific chromatography method (different column chemistry, technique) will be used to achieve further metabolite separation, and thus confirm the identity of the metabolite. The HILIC-HRMS method developed for biomarkers identification could be employed for that purpose.

As a single analytical approach is not enough to cover the entire metabolome, under the generic chromatographic and mass spectrometry conditions used, 119 metabolites contained in the MSMLS™ library were not detected. These include some very highly phosphorylated molecules such as adenosine 5'-triphosphate, cytidine 5'-triphosphate, guanosine 5'-triphosphate, requiring different column chemistry or the addition of ion-pairing buffer to enhance their retention¹⁸.

The integration of multi-approach including both RPLC and HILIC chromatography are necessary to circumvent this issue.

Application to Remote Ischemic Pre-Conditioning

Cardiac remote ischemic preconditioning (RIPC) is a phenomenon whereby transient non-lethal episodes of ischemia applied to a tissue remote from the heart, can protect the myocardium from ischemia-reperfusion (I/R) injury^{19,20}. Though, the actual identities of circulating cardioprotective mediators remain largely unknown.

The rat RIPC cohort was previously exploited in a targeted quantitative metabolomics approach using the Biocrates Absolute IDQ® p180 kit (Biocrates Life sciences AG, Innsbruck, Austria), using completely different extraction protocol (kit user manual), analytical strategy (LC-MS/MS) and platform (QTRAP 5500 mass spectrometer, AB Sciex)¹². The Biocrates Absolute IDQ® p180 kit used in the targeted metabolomics approach enables the quantification of 188 different endogenous molecules. As shown on the Venn diagram in Figure S-2, both libraries have 34 metabolites in common, principally amino acids and biogenic amines. The targeted metabolomics approach led to 124 molecules accurately quantified. Six metabolites (ornithine, glycine, kynurenine, spermine, carnosine and serotonin) were the most important variables for the significant differentiation between RIPC and control groups in the targeted metabolomics study¹². Three of them were confirmed in a human cohort of 50 patients subjected to RIPC, i.e., ornithine, glycine, and kynurenine.

The use of the same sample cohort in the current research had several purposes: (1) to validate the developed workflow, (2) to test the in-house database and (3) to see whether similar biomarkers can be obtained using different analytical platforms and technologies.

Sequence validity

Instrument stability and the validity of the RIPC sequence before further processing was verified as described above. Chromatograms alignment illustrates system stability and chromatographic reproducibility as shown in Figure S-3. PCA was employed to provide an overview of the data structure and ensure clustering of the QCs (Figure S-4). As illustrated in Figure S-4, QC samples are clustered at the center of the PCA, guaranteeing that further discrimination observed would reflect metabolome modification upon RIPC and not analytical variability.

Biomarkers selection from the multivariate and univariate analysis

Data processing produced a number of detected features (ions) aligned by retention time, accurate mass, and peak area. The data matrix was filtered out and this led first to the identification of 147 and 100 ions in positive and negative modes, respectively. After manual curation and removal of duplicate, 104 ions in positive and 68 ions in negative mode were kept in the final data table for multivariate analysis. After PCA, a supervised multivariate OPLS analysis was performed. The OPLS score plot (Figure S-5) displayed a first class separation in positive ionization mode. In fact, control samples (C) have coordinates opposed to the RPC samples (T). The OPLS fit parameters were as follow: $R^2X(\text{cum}) = 0.439$, $R^2Y(\text{cum}) = 0.987$ and $Q^2(\text{cum}) = 0.586$, indicating that most of the variance between the status of the animals (C vs. T) is explained by the model and that the model has prediction ability. This approach clearly distinguished C from T groups. To further specify metabolites associated to the groups' separation, an S-plot was generated and highlighted several potential biomarkers (Figure 2). The examination of the S-plot was combined to jack-knife confidence intervals (as displayed on the loading column plot). This was meant to foresee metabolites with high statistical reliability. Reliable metabolites presenting also variable importance in the projection (VIP) values higher than 1 were subsequently selected as potential biomarkers. Figure S-6, presents in details the selection process of the potential biomarkers.

At this stage, 7 potential biomarkers were revealed, i.e., 5-hydroxyindoleacetate, glycine, kynurenine, ornithine, Marker A (adenosine 5'-monophosphate), Marker B (Aspartate), and Marker C (xanthosine). Univariate statistics were conducted to test the significance of the discriminant biomarkers in C and T samples. The distribution was significantly different from a normal distribution for L-ornithine and Marker C; results for the Kolmogorov-Smirnov test are presented in Table 1. Levene's test for equality of variances indicated unequal variances in C and T samples, for 5-hydroxyindoleacetate, ornithine, Marker A, and Marker C (Table 1). Unequal variance two-tailed t-test was applied to all 7 putative biomarkers, in preference to the Student's t-test and the Mann-Whitney test²¹. Six metabolites out of the seven putative biomarkers were therefore found statistically significant, namely, glycine, kynurenine, ornithine, Marker A, Marker B, and Marker C (Table 1).

Confirmation of the biomarkers was achieved by comparing experimental spectra obtained in rat plasma to authentic metabolite spectra available in the in-house library. This succeeded in perfect matches. Figure 3 provides confirmation of kynurenine and Figure S-7 presents the confirmation data for glycine and ornithine.

Additionally, a HILIC-HRMS method was developed to further ascertain the six biomarkers reported in the present investigation (Figure S-8, Table 2). As could be depicted from Table 2, fold changes are similar between RPLC and HILIC-HRMS methodologies; although slightly lower in the HILIC-HRMS experiments. We attribute this slight difference in fold to the fact that the HILIC experiments

were performed a year after the originally RPLC experiments. The remaining rats plasma stored at -80°C were thawed once more to be re-extracted for confirmation of the biomarkers in HILIC-HRMS.

Table 2 summarizes the characteristics of the identified and statistically significant biomarkers associated to the RIPC investigation. The biomarkers reported in this research were allocated to identification level 1, according to the current MSI reporting standards. From data acquired in negative ionization mode, 68 ions were identified, but none of them were found discriminant and statistically significant between RIPC and control groups.

The results obtained corroborate with findings from the LC-MS/MS targeted metabolomics study previously performed¹². The 3 main biomarkers identified by the latter study (glycine, kynurenine, ornithine), and confirmed in rats and human, were found in the current work. In both research, RIPC was found to be associated with a plasmatic decrease in ornithine and increase in kynurenine and glycine concentrations in rats. Spermine and carnosine were not found in the current investigation, while serotonin was detected but did not exhibit high statistically reliability. As shown on Figure S-5, serotonin did not pass our selection filters; it had a jack-knife confidence interval through zero, a VIP inferior to 1, and was not statistically significant ($p>0.05$) following unequal variance two-tailed t-test.

3 additional cardioprotective metabolites were further evidenced in the present work, i.e., Marker A, Marker B and Marker C; these encouraging findings attests the efficiency of the pipeline presented here, and its ability to identify new biomarkers of clinical interest. Scientific reproducibility in LC-MS based-metabolomics has not often been reported, especially when using different approaches (targeted vs. non targeted), instrument-dependent variability with different MS instrumentations (AB Sciex QTRAP 5500 MS vs. Thermo Fisher Scientific Q Exactive™ HRMS), different sample treatments (MeOH with ammonium acetate followed by derivatization vs. ice-cold MeOH extraction), and different laboratories and researchers.

CONCLUSIONS

The analytical strategy reported here performs adequately and the workflow proved its applicability for metabolomics investigations. It allows a consistent exploration of metabolic signatures, with identification and confirmation of biomarkers in a single experimental sequence, using an in-house metabolites library containing accurate m/z values, retention times, isotopic pattern, full MS and MS/MS spectra. Importantly, the three main biomarkers identified by a previous quantitative MS/MS targeted metabolomics investigation were found in the current investigation, and we further identified three additional potentially cardio protective metabolites. The six biomarkers found in the study were confirmed by RPLC and HILIC-HRMS.

Altogether, these findings prove that scientific reproducibility in metabolomics can be achieved successfully. Further perspectives include the application of the current pipeline to investigations with a greater samples cohort.

REFERENCES

- (1) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; et al. *Nucleic Acids Research* **2007**, 3, D521–D526.
- (2) Want, E.; Nordstrom, A.; Morita, H.; Siuzdak, G. *J. Proteome Res.* **2007**, 6, 459.
- (3) Antignac, J. P.; Courant, F.; Pinel, G.; Bichon, E.; Monteau, F.; Elliott, C.; et al. *TrAC* **2011**, 30(2), 292–301.
- (4) Dettmer, K.; Aronov, P.A.; Hammock, B.D. *Mass Spectrom. Rev.* **2007**, 26(1): 51–78.
- (5) FDA, Guidelines on General Principles of Process Validation, **1987**.
- (6) US FDA, Guidance for Industry (Draft): Analytical Procedures and Methods Validation: Chemistry, Manufacturing, and Controls and Documentation, **2000**.
- (7) US FDA, Guidance for Industry, Bioanalytical Method Validation, **2001**.
- (8) ICH Q2A, Validation of Analytical Procedures: Text and Methodology Q2(R1), Geneva, 1996, in 2005 incorporated in Q2(R1), **2005**.
- (9) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A. et al. *Metabolomics* **2007**, 3(3), 211–221.
- (10) Goodacre, R.; Broadhurst, D.; Smilde, A.K.; Kristal, B.S.; Baker, J.D.; Beger, R. et al. *Metabolomics* **2007**, 3:231–241.
- (11) Naz, S.; Vallejo, M.; García, A.; Barbas, C. *Journal of Chromatography A* **2014**, 1353, 99–105.
- (12) Chao de la Barca, J.M.; Bakhta, O.; Kalakech, H.; Simard, G.; Tamareille, S.; Catros, V.; Callebort, J. et al. Accepted for publication in *Journal of American Heart Association*; Manuscript number JAHA/2016/003891R2.

- (13) Kalakech, H.; Hibert, P.; Prunier-Mirebeau, D.; Tamareille, S.; Letournel, F.; Macchi, L.; Pinet, F.; Furber, A.; Prunier, F. *PLoS One* **2014**, 9:e107950, DOI: 10.1371/journal.pone.0107950.
- (14) Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J. et al. *Nature Protocol* **2010**, 5, 1005–1018.
- (15) Vessman, J.; Stefan, R.I.; Van Staden, J.F.; Danzer, K.; Lindner, W.; Burns, D.T.; et al. *Pure Appl.Chem.*, **2001**, 73 (8)1381–1386.
- (16) van der Kloet, F. M. ; Bobeldijk, I.; Verheij, E. R.; Jellema, R. H. *J. Proteome Res.* **2009**, 8, 5132–5141.
- (17) Eriksson, L.; Trygg, J.; Svante, W. *J. Chemometrics* **2008**, 22(11–12), 594–600.
- (18) Zhou, L.; Xue, X.F.; Zhou, J.H.; Li, Y.; Zhao, J.; Wu, L.M. *J. Agric. Food Chem.* **2012**, 60, 8994–8999.
- (19) Hausenloy, D.J. ; Yellon, D.M. *Nature reviews. Cardiology.* **2011**; 8, 619-629.
- (20) Heusch, G. *Lancet.* **2013**, 381,166-175.
- (21) Ruxton, G.D. *Behavioral Ecology*, **2006**, DOI:10.1093/beheco/ark016.

ACKNOWLEDGMENTS

Research fundings from Université d'Angers, Angers University Hospital, Région Pays de la Loire and Angers Loire Métropole are gratefully acknowledged. The research was performed under the auspice of the PREMMi project in the MitoVasc Institute. We acknowledge fruitful discussions from the MitoLab team members.

CONFLICT OF INTEREST DISCLOSURE

The authors declare no competing financial interest.

ASSOCIATED CONTENT

Supporting information

Further information on the experimental section, figures presenting system stability, PCA, OPLS DA models, Venn diagram of common metabolites between the in-house library and the Biocrates kit, detailed process of biomarker selection, biomarker confirmation, permutation test for final model; and further tables illustrating method performances and list of in-house database metabolites.

AUTHOR INFORMATION:

Corresponding author

* Email: judith.kouassinzoughet@univ-angers.fr

Tel: +33 (0) 244688412 Fax: +33 (0) 244688433

Figure captions

Figure 1: Schematic diagram of the experimental design

Figure 2: S plot for biomarkers selection. Marker A: adenosine 5'-monophosphate, Marker B: Aspartate, Marker C: Xanthosine.

Figure 3: Confirmation of Kynurenine. a) QC dilutions curve (dilutions $\frac{1}{4}$, $\frac{1}{2}$, 1 – top right panel), isotopic pattern (library vs. experimental – top left panel) and compound peak (bottom panel); b) MS2 spectra library vs. experimental.

Figure 1

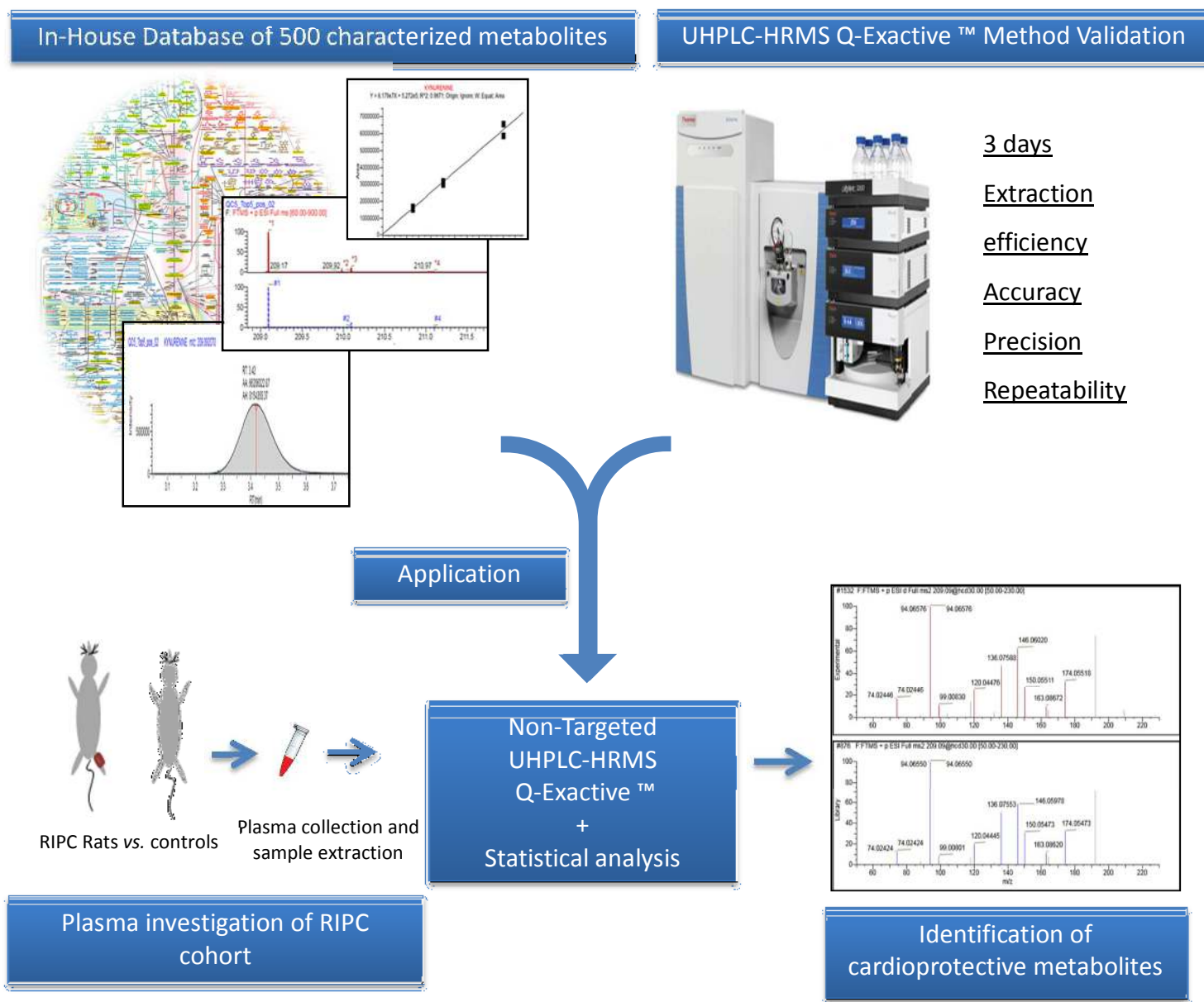


Figure 2

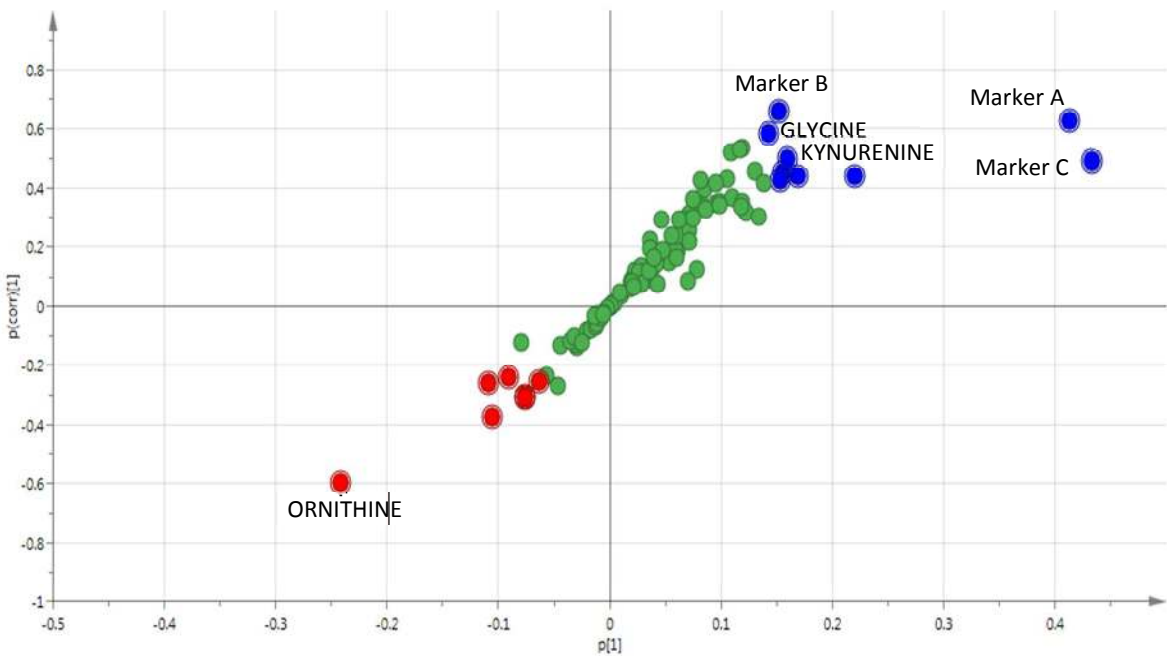
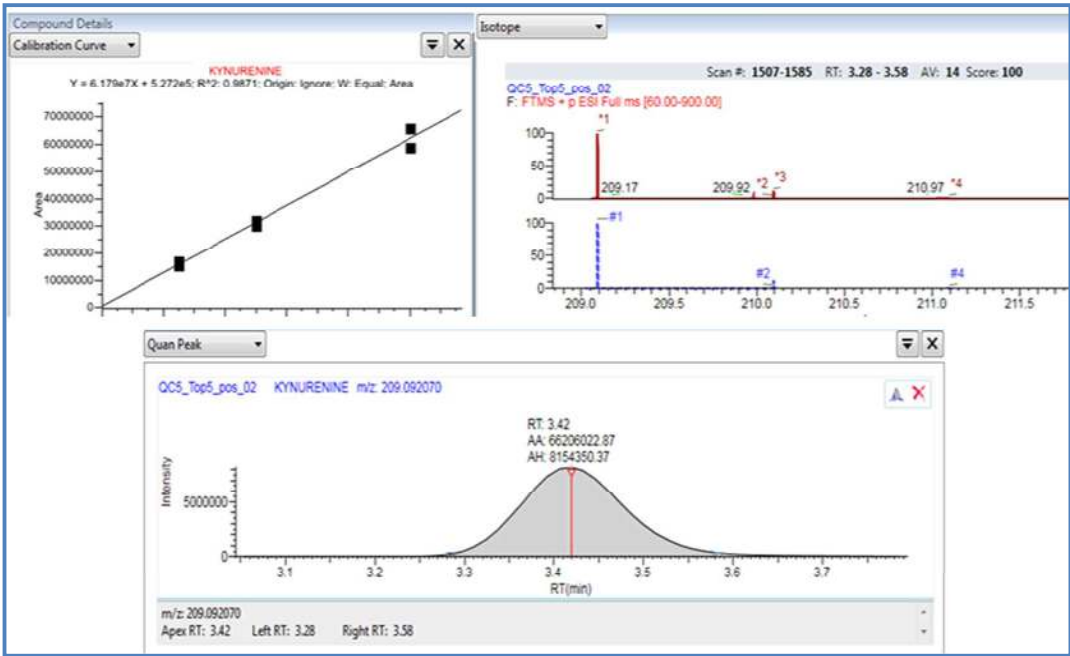


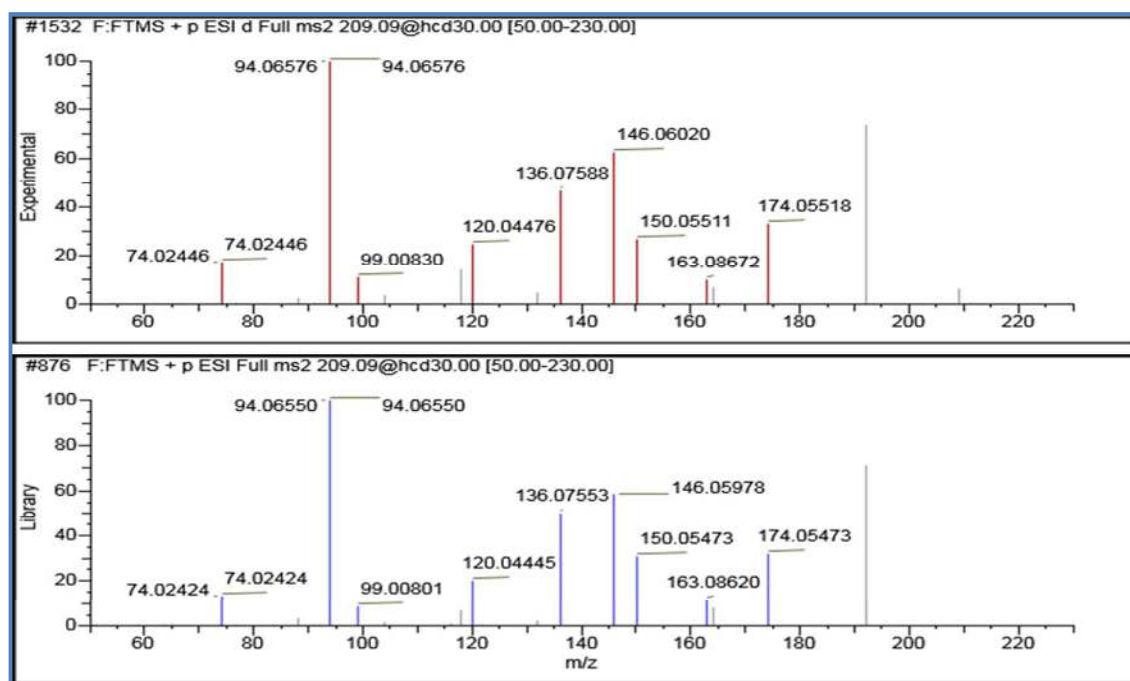
Figure 3

a)



ACCEPTED PAPER

b)



Table

Table 1: Results for the Kolmogorov-Smirnov, Levene’s test, and Unequal variance two-tailed t-test

	Test of Normality - Kolmogorov-Smirnov	
	C samples	T samples
5-hydroxyindoleacetate	D (10) = 0.180, p>0.05	D (10) = 0.145, p>0.05
Glycine	D (10) = 0.133, p>0.05	D (10) = 0.144, p>0.05
Kynurenine	D (10) = 0.195, p>0.05	D (10) = 0.180, p>0.05
Ornithine	D (10) = 0.27, p<0.05	D (10) = 0.285, p<0.05
Marker A (adenosine 5'-monophosphate)	D (10) = 0.226, p>0.05	D (10) = 0.244, p>0.05
Marker B (Aspartate)	D (10) = 0.140, p>0.05	D (10) = 0.184, p>0.05
Marker C (xanthosine)	D (10) = 0.442, p<0.05	D (10) = 0.195, p>0.05
	Levene’s test for equality of variances	
	F(1,18)= 7.985, p<0.05	
5-hydroxyindoleacetate	F(1,18)= 0.016, p>0.05	
Glycine	F(1,18)= 2.176, p>0.05	
Kynurenine	F(1,18)= 6.223, p<0.05	
Ornithine	F(1,18)= 17.22, p<0.01	
Marker A (adenosine 5'-monophosphate)	F(1,18)= 0.396, p>0.05	
Marker B (Aspartate)	F(1,18)= 5.958, p<0.05	
Marker C (xanthosine)		
	Unequal variance two-tailed t-test	
	t(18) = -1.52, p>0.05	
5-hydroxyindoleacetate	t(18) = -2.90, p<0.05	
Glycine	t(18) = -2.31, p<0.05	
Kynurenine	t(18) = 2.57, p<0.05	
Ornithine	t(18) = -2.56, p<0.05	
Marker A (adenosine 5'-monophosphate)	t(18) = -3.74, p<0.01	
Marker B (Aspartate)	t(18) = -2.23, p<0.05	
Marker C (xanthosine)		

Table 2: Characteristics of the identified and statistically significant biomarkers associated to the RIPC investigation.

Metabolites	Adduct	Accurate mass	Up- or Down-regulated	CV%	p^b	Fold change RPLC ^a	Fold change HILIC ^a	Retention time (min) RPLC	Retention time (min) HILIC
Marker A* (adenosine 5'-monophosphate)	[M+H] ⁺	348.07036	Up	7.87	0.030	4.58	4.00	2.39	9.28
Marker B ** (aspartate)	[M+H] ⁺	134.04478	Up	6.88	0.002	1.18	1.17	1.30	8.62
Marker C * (xanthosine)	[M+H] ⁺	285.08296	Up	10.18	0.048	5.89	5.00	6.32	4.94
GLYCINE *	[M+H] ⁺	76.03931	Up	3.14	0.010	1.16	1.08	1.15	7.98
KYNURENINE *	[M+H] ⁺	209.09207	Up	9.31	0.033	1.23	1.10	3.42	6.59
ORNITHINE *	[M+H] ⁺	133.09715	Down	6.55	0.026	-1.62	-1.21	0.89	10.26

^a The fold change was calculated as the ratio of RIPC group (T) to control group (C); ^b P -values were calculated from an unequal variance two-tailed t-test; * $p < 0.05$ ** $p < 0.01$.

ABSTRACT GRAPHIC

