Electronic Supplementary Material (ESI) for Analyst.

Analysis of multi-source metabolomic data using Joint and Individual Variation Explained (JIVE)

Julia Kuligowski[a], David Pérez-Guaita[b], Ángel Sánchez-Illana[a], Zacarías León[c], Miguel de la Guardia[b], Máximo Vento[a,d], Eric F. Lock[e], Guillermo Quintás[f,c]*

[a]Neonatal Research Centre, Health Research Institute La Fe, Valencia, Spain

[b]Department of Analytical Chemistry, University of Valencia, Burjasot, Spain

[c]Analytical Unit, Health Research Institute La Fe, Valencia, Spain

[d]Division of Neonatology, University & Polytechnic Hospital La Fe, Valencia, Spain

[e]Division of Biostatistics, University of Minnesota, Minneapolis, USA

[f]Leitat Technological Center, Bio In Vitro Division, Valencia, Spain

Table of contents

1. Experimental procedures

1.1. Example 1: Plasma collection study

Chemicals and internal standards

All solvents were of LC-MS grade and were purchased from Scharlau (Barcelona, Spain). Ultra-pure water was generated with a Milli-Q water purification system from Merck Millipore (Darmstadt, Germany). The internal standards used for LC-TOFMS were: Deuterated DL-Phenylalanine-D_5, Methionine-D_3, Betaine-D_{11} and PGF2α-D_4 (Cambridge Isotopes Laboratory Inc., Andover, MA, USA), 98% purity. Formic acid (≥95%) was obtained from Sigma-Aldrich Química SA (Madrid, Spain).

Sample preparation

Two blood samples were collected from a volunteer in heparin and, alternatively, in EDTA-K3 vacutainer tubes. Blood samples were centrifuged at 1800 x g during 10 min at 20°C and the plasma layer was removed and transferred to Eppendorf vials that were then centrifuged at 2000 x g for 5 min at 20°C to remove blood platelets. Then, EDTA-K3 and heparinized plasmas were withdrawn and stored at -80°C until analysis. Blank

samples prepared by adding 3 mL of ultrapure H_2O to heparin and EDTA-K3 collection tubes were pretreated as blood samples.

For processing, both plasma and blank samples were kept in an ice bath until analysis to prevent degradation. After thawing, the plasma sample was homogenized on a vortex mixer during 10 s. Then, 300 µL of ice-cold CH_3CN, CH_3OH or acetone were added to 100 µL of plasma and blank samples for protein precipitation. For each plasma sample and protein precipitation, a total of six replicates were prepared. After centrifugation at 10000 x g for 15 min at 4°C, 100 µL of the plasma samples were transferred to clean Eppendorf vials that were dried in a speedvac concentrator at 70°C. Then, for UPLC-ESI(-)TOFMS analysis, two of the dried aliquots of each plasma were redissolved in 100 µL of an internal standard solution containing 5 µM DL-Phenylalanine-D_5, Methionine-D_3, Betaine-D_{11} and PGF2α-D_4 in (H_2O:CH_3CN, 0.05% HCOOH v/v) (95:5). The quality control (QC) sample was prepared by mixing 10 µL of each sample solution from the 96-well plate prior to UPLC-TOF-MS analysis.

For ATR-FTIR analysis the dried residues were redissolved in (CH_3OH:acetone) (1:1) to facilitate solvent evaporation in the IR interface.

UPLC-TOF-MS analysis

Chromatographic analysis of the plasma samples was performed on an Acquity UPLC chromatograph using an Acquity UPLC HSS T3 (100 x 2.1 mm, 1.8 µm) analytical column from Waters (Wexford, Ireland). Autosampler and column temperatures were set to 4°C and 40°C, respectively and the injection volume was 5 µL. A gradient elution with a total run time of 14 min was performed at a flow rate of 500 µL min^{-1} as follows: initial conditions of 100% of mobile phase A (H_2O (0.1% v/v HCOOH)) were kept for 1 min, followed by three linear gradients from 0% to 15% of mobile phase B (CH_3CN (0.1% v/v HCOOH)) in 2 min; from 15% to 50% B in 3 min; and from 50% to 95% B in 3 min; 95% B was held for 3 min and finally, a 0.5 min gradient was used to return to the initial conditions, which were held for 1.5 min. Full scan MS data from 50 to 1000 mass to charge ratio (m/z) with a scan time of 0.1 s was collected on a quadrupole time of flight (QTOF) SYNAPT HDMS spectrometer (Waters, Manchester, UK) in the TOF MS mode. The following electrospray ionization parameters were selected in both, the ESI(+) and ESI(-) with extended dynamic range: capillary and cone voltages were set at 3.2 kV and 20 V; source and desolvation temperatures were set at 120°C and 380°C, respectively; flow rates of cone and nebulization gases were set at 50 and 800 L/h, respectively.

Sample acquisition was randomized to avoid bias effects of instrument drifts during the LC batch. Solvent blank samples were analyzed at the end of the batch to control background and cross-contamination and the QC sample was analyzed every 6 samples to monitor instrumental response. A set of seven QC samples was injected at the beginning of the batch for column conditioning. The sample ordering in ESI(+) and ESI(-) was replicated to facilitate the identification of instrumental batch effects. Between ESI(+) and ESI(-) sample analysis, the ESI/MS detector inlet interface was cleaned and the MS recalibrated.

ATR-FTIR analysis

Infrared spectra were obtained on a Bruker (Bremen, Germany) IFS 66/v FTIR spectrometer equipped with a liquid nitrogen-refrigerated mercury–cadmium–telluride detector, a vacuum system, and a dry air purged sample compartment. Measurements were made using an ATR DuraSampleIR accessory with a nine reflections diamond/ZnSe DuraDisk (Smiths Detection Inc., Warrington, UK). For sample analysis, 1 µL of plasma was deposited in the center of the ATR crystal and then dried at room temperature in approximately 1 min. After that, the spectrum of each sample was collected co-adding 300 scans in the range between 4000 and 600 cm^{-1} at a resolution of 4 cm^{-1} using a zero filling factor of 2 and a spectrum of the ATR crystal as background. In order to evaluate the repeatability of the measurement, each sample was dried and measured by triplicate. After each measurement, the ATR surface was thoroughly cleaned using H_2O, CH_3CN and the solvent used for protein precipitation. Spectral acquisition order was randomized to avoid biased results due to instrumental effects. Spectra were baseline corrected using a polynomial function of order 1 and 3 for the fingerprint and the CH_3 regions respectively.

LC-TOF-MS data processing

Centroid raw LC-TOF-MS data (.raw files) was converted into .netCDF format using DataBridge (Waters) before generating peak tables using XCMS software (http://metlin.scripps.edu/xcms/). The centWave method was used for peak detection with the following parameters: ppm=20, peakwidth=(2, 25), snthresh=5. A minimum difference in m/z of 5 mDa was selected for peaks with overlapping RTs. Intensity weighted m/z values of each feature were calculated using the wMean function. Peak limits used for integration were found through descent on the Mexican hat filtered data. Peak grouping was carried out using the 'nearest' method, using the following parameters, mzVsRT=3 and RT and m/z tolerances of 10 s and 5 mDa, respectively. After

peak grouping, the fillPeaks method with the default parameters was applied to fill missing peak data.

The obtained peak tables were imported into MATLAB (Mathworks Inc., Natick, MA, USA) for data analysis. A total of 3320 features were initially present in the ESI(+) data source after peak detection, chromatographic de-convolution and peak alignment of the entire chromatographic batch. Blank samples were used to identify a total of 2445 features arising from source contaminants and other sample components originating from e.g. tubes, solvent impurities, anticoagulant that were removed. Then, variables showing %RSD>25 in the QC samples were removed leaving a total of 339 features. When the plasma samples were analysed using ESI(-), the number of initially detected features was lower (2045) and was reduced down to 437, after elimination variables detected in blanks (948) and those showing %RSD>25 in the QC samples (660).

The mean intensities of the internal standards (ISs) DL-Phenylalanine-D_5, (RT: 2.71 min, ESI(+)/-) Methionine-D_3 (RT: 0.70 min, ESI(+)/-), Betaine-D_{11} (RT: 0.49 min, ESI(+)) and PGF2α (RT: 6.00 min, ESI(-)) were used to monitor instrument performance. Whereas the intensities of the ISs were comparable among QC and plasma samples, higher intensities of Methionine-D_3 and Betaine-D_{11} were found in blanks in agreement to lower levels of ion suppression. Using ESI(-), the intensity of Methionine-D_3 in samples was very low and it was only detected in blanks. No IS correction was performed to keep the potential effect of ionic suppression as a discriminant variable among the six different plasma collection and protein precipitation procedures.

1.2 Example 2: Plasma and urine simultaneous analysis (PAUSA) study

Chemicals and internal standards

LC-MS grade methanol (CH_3OH), water and acetonitrile (CH_3CN) were purchased from Fisher Scientific (Loughborough, U.K.). Formic acid (FA) (≥95% purity) was obtained from Sigma-Aldrich Química SA (Madrid, Spain). The internal standards (IS) were: Deuterated DL-Phenylalanine-D_5, ($PheD_5$), Methionine-D_3 ($MetD_3$), Betaine-D_{11} (BetD11) and 8-iso Prostaglandin F2α-D_4 (PGF2α-D_4) (98% purity) were purchased from Cambridge Isotopes Laboratory Inc. Reserpine (Res) and Leucine Enkephalin (LeuEnk) (>95% purity) from Sigma-Aldrich Quimica SA (Madrid, Spain). An IS solution was prepared by mixing 2 mL of water (0.1% FA, v/v) containing 2.5 µM $PheD_5$, 6 µM $MetD_3$, 2.5 µM PGF2α-D_4 and 12 µM $BetD_{11}$ and 2 mL of water (10% ACN, v/v) containing 1.6 µM Res and 1.8 µM LeuEnk.

Sample preparation

Blood and urine samples were collected from four volunteers before and 4 h after consumption of green tea. Participants were instructed to refrain from tea consumption three days prior to sampling.

Blood samples were collected in 3 mL EDTA-K3 vacutainer tubes and then centrifuged at 1400 x g during 10 min at 20°C. Then, 1 mL of the plasma layer was removed from each sample and transferred to a 10 mL polypropylene tube that was centrifuged at 2500 x g for 15 min at 20°C to remove blood platelets. A plasma blank sample was prepared by adding 3 mL of water to an EDTA-K3 Vacutainer tube applying the same treatment as described for blood extracts.

In the case of urine samples, the volunteers were instructed to collect midstream specimens at the scheduled times in sterile containers. Following collection, samples were centrifuged at 2000 x g during 15 min at 4°C to precipitate sediment and then supernatants were transferred to microcentrifuge tubes. A urine blank sample was prepared by adding 30 mL of water to a sterile container and treated as urine samples. Quality control (QC) solutions prepared by mixing 20 µL from each plasma and urine supernatant underwent the same procedure as the samples.

For sample processing, plasma, urine, QCs and blank samples were kept in an ice bath to prevent degradation. 100 µL of each plasma sample were added to 300 µL of ice-cold CH_3CN for protein precipitation. After centrifugation at 10000 x g for 10 min at 4°C, 100 µL of the supernatants were transferred to clean tubes and dried in a Savant speedvac concentrator (Thermo electron corporation, USA) and the residue was then reconstituted in 100 µL of the IS solution. Finally, the obtained sample solutions were transferred into 350 µL volume 96-well plates for analysis.

50 µL of urine and blank samples were diluted with 100 µL of the IS solution. The obtained urine sample solutions were transferred into 350 µL volume 96-well plates for analysis.

UPLC-TOF-MS analysis

Chromatographic analysis of both plasma and urine samples was performed on an Acquity UPLC chromatograph using an Acquity UPLC HSS T3 (100 x 2.1 mm, 1.8 µm) analytical column from Waters (Wexford, Ireland). UPLC-ESI(+)-TOF-MS instrumental parameters selected in the *Plasma collection* study described above were used for sample

analysis. Again, sample acquisition was randomized to avoid bias effects due to instrument drifts during the LC batch. A blank sample was analyzed every 4 plasma or urine samples to control background and cross-contamination and a QC sample was analyzed after each blank sample. A set of five blank samples and ten QC solutions were analyzed at the beginning of the batch for column conditioning. The random sample ordering in ESI + was mantained in the ESI(-) analysis.

UPLC-TOF-MS data processing

Peak table generation was performed as described for the *Plasma collection* study. For each biofluid and ionization conditions, blank samples were used to identify features arising from source contaminants and other sample components originating from e.g. tubes, solvent impurities or anticoagulant that were removed from the data. Then, variables showing %RSD>25 in the QC samples were removed from further analysis. The intensities of the internal standards (ISs) DL-Phenylalanine-D_5, (RT: 2.71 min) Methionine-D_3 (RT: 0.70 min), Betaine-D_{11} (RT: 0.49 min), Reserpine (RT: 6.02 min) and LeukineEnkephaline (RT: 4.61 min) were used to monitor instrument performance. No IS normalization was performed.