Steps in designing and preparing the spiked-in metabolomics data set

Estimation of physiologic metabolite concentrations

Using preliminary GC-MS data from clinical samples in our laboratory, 12 potentially clinically-relevant metabolites were selected for analysis (alanine (Ala), arginine (Arg), aspartic acid (Asp), glycine (Gly), leucine (Leu), lysine (Lys), proline (Pro), pyroglutamic acid (Pyr), serine (Ser), threonine (Thr), tyrosine (Tyr), and valine (Val)). Concentration data obtained from previous quantitative nuclear magnetic resonance (NMR) studies in our laboratory were used to estimate the target spiked concentration of these 12 metabolites. For each metabolite, the minimum concentration (as an estimate of background physiologic levels) was subtracted from the mean concentration to determine the amount of pure metabolite that would have to be added to standardized serum to achieve the target spiked concentration. Maximum spiked concentration was calculated at 2X the target concentration.

Preparation of serum and metabolite solutions

Standardized pooled human serum was obtained from the National Institute for Standards and Technology (NIST) and stored at -80C. This standard serum was thawed on ice for 1 hour prior to preparation of metabolite solutions. Standard phosphate buffer solution (PBS) was prepared [NaCl/KCl/Na2HPO4/KH2PO4, pH 7.4 ±0.05]. A 10% (v/v) mixture of PBS in standard serum was prepared to use as control. Standard metabolites were obtained from Sigma (St. Louis, MO) and stored according to manufacturers specifications. Metabolite solution 1 was mixed at 10X maximal concentration in standard PBS using exact weights of glycine, serine, leucine, threonine, tyrosine, and aspartic acid. Metabolite solution 2 was similarly mixed at 10X maximal concentration in standard PBS using exact weights of alanine, valine, lysine, proline, arginine, and pyroglutamic acid. Each solution was vortexed at room temperature and at 20oC until all of the metabolites had completely dissolved. For each 10X concentrated metabolite solution, 1 mL was mixed with 9 mL of standard human serum to create serum solutions with goal maximum spiked concentrations. In addition, for each 10X concentrated metabolite solution, 1 mL was mixed with 9 mL of standard PBS to create 1X PBS solutions with goal maximum spiked concentrations.

Preparation of dilution series

For each maximally concentrated spiked metabolite serum, a dilution series in standardized human serum was created by pipetting 50 uL, 42 uL, 33 uL, 25 uL, 17 uL, and 8 uL into each of 6 labelled 2-mL tubes. To each tube, 0 uL, 8 uL, 17 uL, 25 uL, 33 uL, and 42 uL, respectively, of 10% PBS in standard serum solution was added to make the total volume in each tube 50 uL. The concentration factors (CF) for these tubes, relative to the midpoint in the series, were 2.0, 1.67. 1.33, 1.0, 0.67, and 0.33, respectively. This process was repeated to create triplicate samples for each dilution series. Triplicate control samples (CF 0) were created by pipetting 50 uL of 10% PBS in standard serum solution into 2-mL tubes. A parallel process was used to create a dilution series of spiked metabolite solutions in PBS (no serum).

Sample extraction and derivatization

Extraction and derivatization were conducted according to a protocol based on that of A et al.42 Each 50 uL sample underwent deproteination with 300 µL of cold 2:1 methanol:chloroform. The aqueous metabolite fraction was extracted by chloroform separation with centrifuge at 13,300 rpm for 7 minutes. The aqueous fraction was dried overnight by speedvac (VWR Scientific

Products, Radnor, PA). All samples were then derivatized with methoxylamine in pyridine for 3 hours at 37oC with shaking, and then with methyltrimethylsilyltrifluoroacetamide (MSTFA) for 1 hour at 37oC with shaking. Each derivatized sample was then diluted with 500 uL hexane and centrifuged for 4 minutes at 13,200 rpm. GC-MS vials, with glass inserts, were loaded with 200 uL of each derivatized sample.

Gas chromatography-mass spectrometry (GC-MS)

After extraction and derivatization, each sample underwent GC-MS on a Waters Technology machine using a 30 m x 0.25 mm column with a 0.25 μm DB-5MS stationary phase, helium at a flow rate of 1.2 mL/min, and splitless injection of 1 μL of sample. A ramped temperature program was used, starting at 80oC for 8 minutes, increasing from 80 to 320oC by 12oC/min, and held there for 8 minutes. Mass spectrometry used electron ionization with a 70 eV electron beam and a time-of-flight mass analyzer. Scan range is 50-800 m/z at a rate of 6.7 scans per second. Alkane standards were run before, after every 20 samples, and at the end of the batch. Control and dilution series samples were all run, in randomized order, on the same day.

Data processing

Retention index (RI) calibration, peak deconvolution, metabolite identification and quantification were conducted, with a non-targeted approach, using MetaboliteDetector (Version 2.06beta)45 with the following settings: peak threshold: 3.00; minimum peak height: 3.00; bins per scan: 2; width: 1.90; maximum RI window: 20; pure/impure ratio: 0.30; cutoff score 0.70; minimum distance between peaks: 0.50; minimum quality index: 1.00; delta RI for compound matching: 5.0; delta RI for identification: 20; scoring method: RI + spectra; required score for compound matching: 0.85; required reproducibility score: 0.30; maximum peak discriminating index: 100; minimum required signal/noise ratio: 0; and minimum required number of ions: 1. There were no significant differences between standard alkane elution times over the course of the experiment and RI calibration was conducted using the mid-point alkane standard retention times.

The Golm Metabolome database was used as the library for metabolite identification. The main ions produced by independent trimethylsilyl groups (m/z 73 and 146) were excluded prior to metabolite identification and quantification. The serum and PBS samples were quantified in separate data analysis batches.

For each spiked metabolite, the maximally derivatized species was chosen for further analysis, based on the consistency of ion abundancies measured in the three control standard serum samples. Table 1 lists these species, along with their Golm reference parameters (retention index (RI) and select mass ions) and measured parameters (mean RI and measured ions). Non-maximally silyated species of the spiked metabolites were excluded from further analysis. All remaining metabolite features, whether matched to an identification by MetaboliteDetector or not, were included in further analysis as background features of the standard serum or PBS solution. PQN normalization66 was applied to all spectra using R (www.r-project.org).