UPLC/MS Analysis of Human Serum Samples from the Hepatitis B Research Network Cohort using the Biocrates Absolute*IDQ* p400HR Kit

In collaboration with Keyur Patel, University of Toronto

March 14, 2018

Objective: Measure the levels of selected metabolites in 60 human serum samples using the Biocrates Absolute*IDQ* p400 HR kit.

Duke Proteomics Core Facility Contributors: Lisa St. John-Williams (sample preparation, data collection, data analysis, report writing), Will Thompson (study design, scientific oversight, report writing), and Arthur Moseley (scientific oversight, report writing).

Introduction:

The Absolute*IDQ* p400HR assay quantifies over 400 metabolites from eleven analyte groups: amino acids, biogenic amines, acylcarnitines, monosaccharides (hexose), diglycerides, triglycerides, lysophposphatidylcholines, phosphatidylcholines, sphingomyelins, ceremides, and cholesteryl esters. The p400HR kit includes all requisite calibration standards, internal standards, and QC samples. The use of these standards according to the detailed analysis protocol which was validated in Biocrates' lab in Austria assures assay harmonization and standardization within a project, across projects, and across laboratories. Selective analyte detection is accomplished by use of a high-resolution, accurate-mass mass spectrometer with a Q Exactive Plus Orbitrap™ mass analyzer. There are four separate mass spectrometric analyses of each sample. For the analysis of acylcarnitines, monosaccharides (hexose), diglycerides, triglycerides, lysophposphatidylcholines, phosphatidylcholines, sphingomyelins, ceremides, and cholesteryl esters samples are introduced using two different Flow Injection Analysis methods (FIA-MS, **Figure 1 A**). Sample analysis of amino acids and biogenic amines are performed by two UPLC (ultra-high pressure liquid chromatography) methods (one using full scan MS and one using parallel reaction monitoring or PRM) using a reversed phase analytical column for analyte separation (LC-MS, **Figure 1B**).

Figure 1. Schematic of data collection methodologies for Biocrates p400HR kit, including Flow-Injection MS (A) and LC-MS (B).

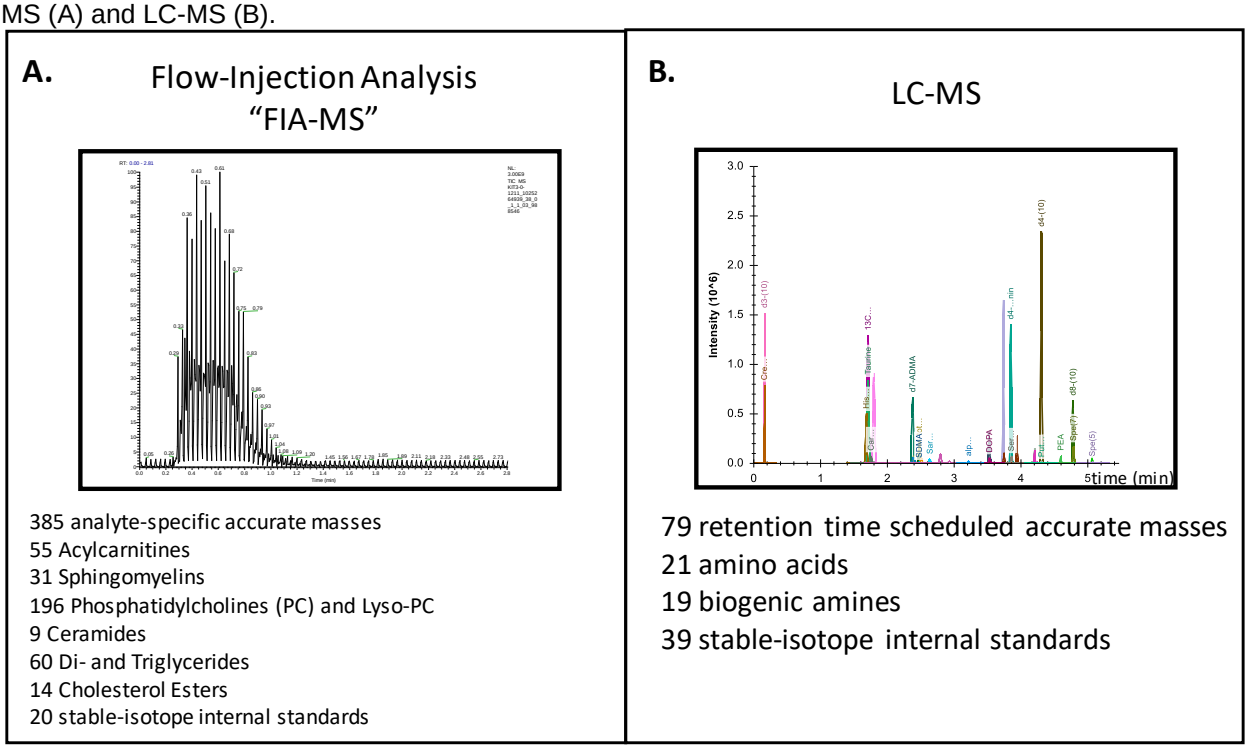
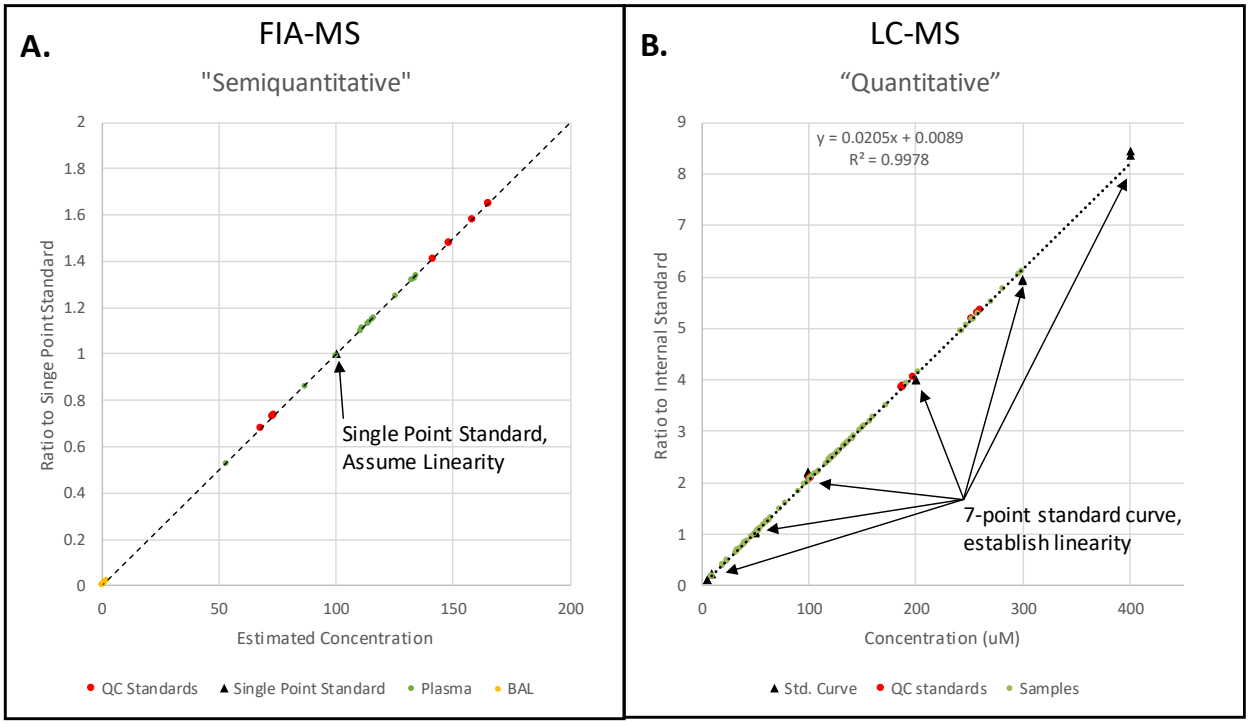


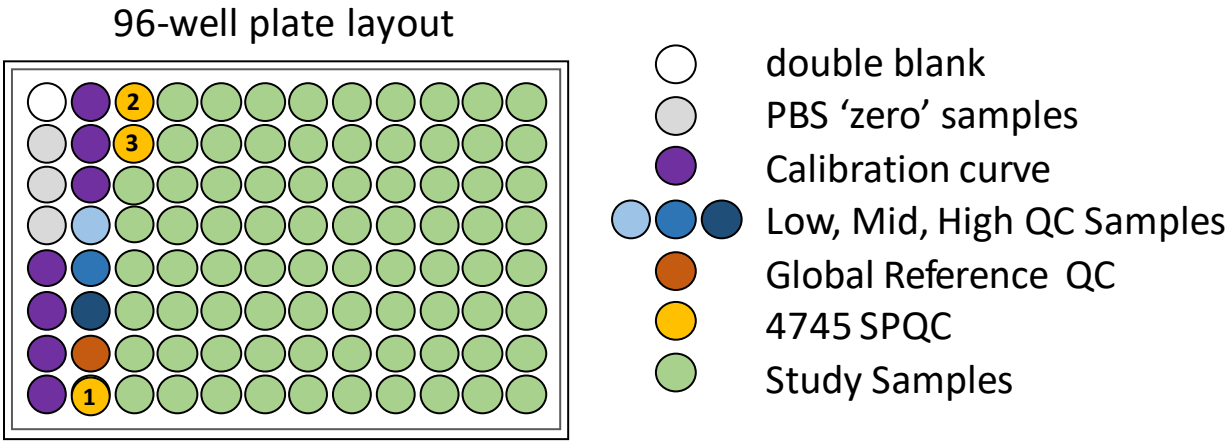
Figure 2 shows a schematic and representative examples of how the single-point calibrator or calibration curve would be used to back-calculate QC and sample concentrations. A single point calibration standard is used for the reproducible quantitation of the acylcarnitines, monosaccharides (hexose), diglycerides, triglycerides, lysophposphatidylcholines, phosphatidylcholines, sphingomyelins, ceremides, and cholesteryl esters. Therefore, the results for these analytes are described as semi-quantitative (schematic shown in **Figure 2A**). The seven calibration standards provided in the Biocrates Absolute*IDQ* kit are used for highly accurate and reproducible quantitation of the amino acids and biogenic amines as shown in **Figure 2B**. Calibration standards were fit with a quadratic regression using $1/x^2$ weighting.

Figure 2. Schematic depicting the quantitative methodologies used in the Biocrates p400HR kit for flow-injection analysis (A) and LC-MS analysis (B).



The samples were prepared in a 96-well plate format using the layout shown in **Figure 3**.

Figure 3. Schematic depicting the 96-well plate layout for the analysis of serum samples (Plates 1 and 2) including: blanks, calibration standards, and QC samples from Biocrates. Two additional QC samples were analyzed: the DPMCF Global Reference QC and the serum study sample pool QC (4745 SPQC).



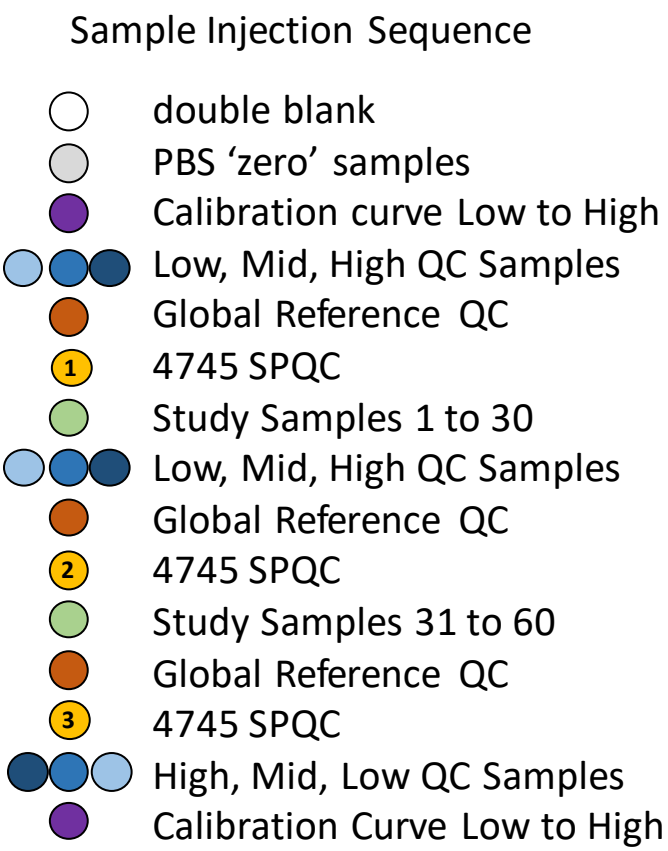
Sample Preparation:

DPMSR Project 4745

Samples were prepared using the Absolute*IDQ*® p400HR kit (Biocrates Innsbruck, Austria) in strict accordance with their detailed protocol. After the addition of 10 µL of the supplied internal standard solution to each well of the 96-well extraction plate, 10 µL of each study sample were added to the appropriate wells. The plate was then dried under a gentle stream of nitrogen. The samples were derivatized with phenyl isothiocyanate then eluted with 5mM ammonium acetate in methanol. Samples were diluted with either water for the UPLC analysis (1:1) or running solvent (a proprietary mixture provided by Biocrates) for flow injection analysis (5:1).

A pool of equal volumes of all 60 serum samples analyzed on the first plate was created (4745 SPQC), aliquoted, and frozen. The pooled sample was prepared and analyzed in the same way as the study samples on both plates. From each plate this sample was injected once before, once during, and once after the study samples in order to measure the performance of the assay across the sample cohort. The analyses of this pool can be used to assess potential batch effects. The order of injection of the samples is shown in **Figure 4**.

Figure 4. Schematic depicting the injection order of the samples for UPLC analysis. For Flow Injection Analysis the injection order is the same except the calibration curves are not analyzed. Note that the Global Reference QC was prepared once and analyzed three times giving a measure of the analytical variability of the MS analyses. The SPQC sample was prepared three times and each extract analyzed once giving a measure of total analytical variability: sample preparation variability and UPLC/MS variability.



Sample Analysis:

UPLC separation of amino acids and biogenic amines was performed using a Waters (Milford, MA) Acquity UPLC with a proprietary column and guard column provided by Biocrates. Analytes were separated using a gradient from 0.2% formic acid in water, to 0.2% formic acid in acetonitrile. Total UPLC analysis time was approximately 6 minutes per sample. Acylcarnitines, monosaccharides (hexose), diglycerides, triglycerides, lysophposphatidylcholines, phosphatidylcholines, sphingomyelins, ceremides, and cholesteryl esters were analyzed by flow injection analysis (FIA) with total analysis time of approximately 3.8 minutes per sample. Using electrospray ionization in positive mode, samples for both UPLC and flow injection analysis were introduced directly into a Thermo Q Exactive Plus mass spectrometer (Thermo) operating in the full scan or parallel reaction monitoring (PRM) mode. Accurate masses (compound-specific) for each analyte and internal standard were collected over the appropriate retention time. The UPLC-MS data were imported into Thermo Xcalibur application QuanBrowser for peak integration, calibration, and concentration calculations. The UPLC-MS data from QuanBrowser and FIA-MS data were analyzed using Biocrates Me*tIDQ*™ software.

Data Return Document Descriptions

A number of documents have been added to the Express Data Repository for Project 4745 including this sample analysis summary. A link to the repository is included here. A description of the data spreadsheets follows the link.
https://discovery.genome.duke.edu/express/resources/4745/.

4745 UPLC p400 Data.xlsx

This is an Excel workbook containing three worksheets. The first worksheet contains the calculated concentration data (µM) acquired in the study for analyte classes Amino Acids and Biogenic Amines. The second row in this worksheet lists the analytes measured and the status for each one. Row 4 lists the lower limit of detection (LOD) for the analytes. The Biocrates-defined lowest calibration standard and highest calibration standard are listed in Rows 5 and 6. The barcode number of the plate is listed in column A. The Sample Bar Code number (column B) is assigned to every sample by the Biocrates Me*tIDQ*™ software. The Sample Identification Number in column D is the unique sample identifier assigned by the Proteomics and Metabolomics Sample Submission System. Column G lists the Customer Sample Identification information which were listed on every sample tube provided for analysis. Table 1 below gives the unique plate barcodes for the LC-MS and Flow-Injection Analysis-MS (FIA-MS) analysis. The results for the 4745 SPQC sample are in rows 68 through 70. Row 72 has the %CV for the 4745 SPQC sample results. The average %CV for the UPLC data for 4745 SPQC is 5.6%, with a range from 0.9 to 28.8%. It is our recommendation that if the inter-day variability of the Study Pool QC sample (as measured by the %CV) for a particular analyte is greater than 25%, then the data for that analyte in the study samples should be flagged for removal from the dataset because of imprecision. Additionally, best practice suggests that analytes which have more than 40% missing values should be flagged for removal from the dataset because of excess missing values or separated for an independent assessment. It should be noted that in a case-control study with severe metabolite dysregulation, a high level of missing values may be observed in some samples. In this case, replacement of missing values with LOD/2 is recommended.

In addition, some scaling to control for batch effects may be needed during statistical analysis; the SPQC sample can provide a valuable measurement to utilize for controlling batch effects. Further investigation of batch effect is discussed below under Principal Components Analysis.

Table 1. Plate Barcodes

Plate Barcodes for UPLC-MS Analysis.	Plate Barcodes for FIA-MS Analysis
1025266578-1 | 1025266582-1	1025266597-2 | 1025266608-2

In the data document ***4745 UPLC p400 Data.xlsx*** the concentration data (µM) are coded as shown below in Table 2 in order to allow presentation of additional information regarding data quality:

Table 2. Key to the Analyte Status Columns Used in the 4745 Data Tables

Valid	Calculated concentrations are based on a standard curve of the analyte listed. (Not all analytes are contained in the calibration standards provided by Biocrates. Therefore, results for some anlaytes will be coded as Semi-Quantitative.)
<Lowest CS: Lowest Calibration Standard>Value>LOD	The value is greater than the LOD but less than the Biocrates-defined lowest calibration standard. These values should be considered reliable except for analytes for which the %CV observed for the pooled sample was greater than 25%.
>Highest CS	Greater than the highest calibration standard
<LOD	Less than the Lower Limit of Detection. The LODs were given by Biocrates, and were not determined using FDA guidelines or a Waters TQ-S mass spectrometer.
Internal Standard out of range	The internal standard peak area for this analyte was outside of the normal range. (Not all analytes have internal standards.)
Semi-Quantitative	Calculated concentration is not based on a calibration curve but on the peak area of an internal standard or that of a structurally similar analyte. This data is valid but is semi-quantitative instead of fully quantitative.
No Interception or NA	No peak was detected in the chromatogram at the appropriate retention time for this analyte. For statistical analyses apply the same value used for <LOD.

The second worksheet in the workbook contains the study sample data with the Status columns removed.

The third worksheet in the workbook contains a statistical summarization of the data from the Biocrates Met*IDQ*™ application. The Analyte Statistics worksheet lists these summary statistics (Min µM, Max µM, Mean µM, Median µM, 25th Percentile µM, 75th Percentile µM, STD µM, MAD µM, Skewness, Kurtosis, CV [%], CVRobust [%]) by metabolite for all of the samples. The number of results (e.g. samples with usable data) included for each statistical calculation is indicated by "n" in Column C. These results are compiled for the entirety of the dataset, in essence showing the biological variance in the analytes measured.

4745 UPLC p400 QC Data.xlsx

This workbook contains the Biocrates plasma QC concentration data for three levels of QCs for analyte classes amino acids and biogenic amines. The measurements for the QC samples are provided in order

to confirm that quantification of the metabolites across the wide dynamic range performed in this analysis is generally accurate and reproducible. QC values showed excellent reproducibility (Table 3).

Table 3. Reproducibility of UPLC Biocrates QC Sample Analyses

QC Level	Average % CV	Min %CV	Max %CV
Low	4.5	0.0	25.2
Mid	3.2	0.1	15.2
High	3.4	0.4	34.1

4745 FIA p400 Data.xlsx

This is an Excel workbook containing three worksheets. The first worksheet contains the calculated concentration data (µM) acquired in the study for analyte classes acylcarnitines, monosaccharides (hexose), diglycerides, triglycerides, lysophposphatidylcholines, phosphatidylcholines, sphingomyelins, ceremides, and cholesteryl esters. The second row in this worksheet lists the analytes measured and the status for each one. Rows 4 and 5 list the lower limit of detection (LOD) for the analyte in each analytical run. The Biocrates-defined lowest calibration standard and highest calibration standard are listed in Rows 6 and 7 if applicable. The barcode number of the plate on which each sample was analyzed is listed in column A. The Sample Bar Code number (column B) is assigned to every sample by the Biocrates MetIDQ software. The Sample Identification Number in column D is the unique sample identifier assigned by the Proteomics and Metabolomics Sample Submission System. Column G lists The Customer Sample Identification information which were listed on every sample tube provided for analysis. Table 1 above gives the unique plate barcode for the LC-MS and Flow-Injection Analysis-MS (FIA-MS) plates. The results for the 4745 SPQC sample are in rows 68 through 70. Row 72 has the %CV for the 4745 SPQC sample results. The average %CV for the FIA data for 4745 SPQC is 14.6%, with a range from 0.5 to 134.3%. It is our recommendation that if the inter-day variability of the Study Pool QC sample (as measured by the %CV) for a particular analyte is greater than 25%, then the data for that analyte in the study samples should be flagged for removal from the dataset because of imprecision. Additionally, best practice suggests that analytes which have more than 40% missing values should be flagged for removal from the dataset because of excess missing values or separated for an independent assessment. . It should be noted that in a case-control study with severe metabolite dysregulation, a high level of missing values may be observed in some samples. In this case, replacement of missing study sample values with LOD/2 is recommended.

In the data document ***4745 FIA p400 Data.xlsx*** the concentration data (µM) are coded as shown above in Table 2 in order to allow presentation of additional information regarding data quality.

The second worksheet in the workbook contains the study sample data with the Status columns removed.

The third worksheet in the workbook contains a statistical summarization of the data from the Biocrates Met*IDQ*™ application. The Analyte Statistics worksheet lists these summary statistics (Min µM, Max µM, Mean µM, Median µM, 25th Percentile µM, 75th Percentile µM, STD µM, MAD µM, Skewness, Kurtosis, CV [%], CVRobust [%]) by metabolite for all of the samples. The number of results (e.g. samples with usable data) included for each statistical calculation is indicated by “n” in Column C.

4745 FIA p400 QC Data.xlsx

This workbook contains the Biocrates plasma QC concentration data for three levels of QCs for analyte classes acylcarnitines, glycerophospholipids, and sphingolipids. QC concentrations are not provided for all the analytes detected, only for the subset of analytes within each class which have reference values available from Biocrates. The measurements for the QC samples are provided in order to confirm that quantification of the metabolites across the wide dynamic range performed in this analysis is accurate and reproducible. These Biocrates QC values showed excellent reproducibility (Table 4).

Table 4. Reproducibility of FIA Biocrates QC Sample Analyses

QC Level	Average % CV	Min %CV	Max %CV
Low	3.8	0.1	14.3
Mid	3.6	0.3	14.6
High	3.2	0.2	13.9

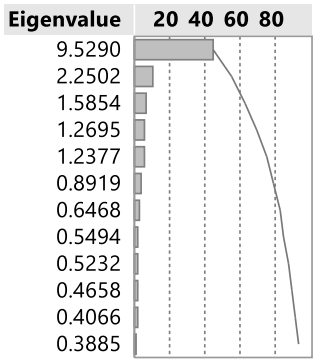
Principal Components Analysis (PCA) Plots

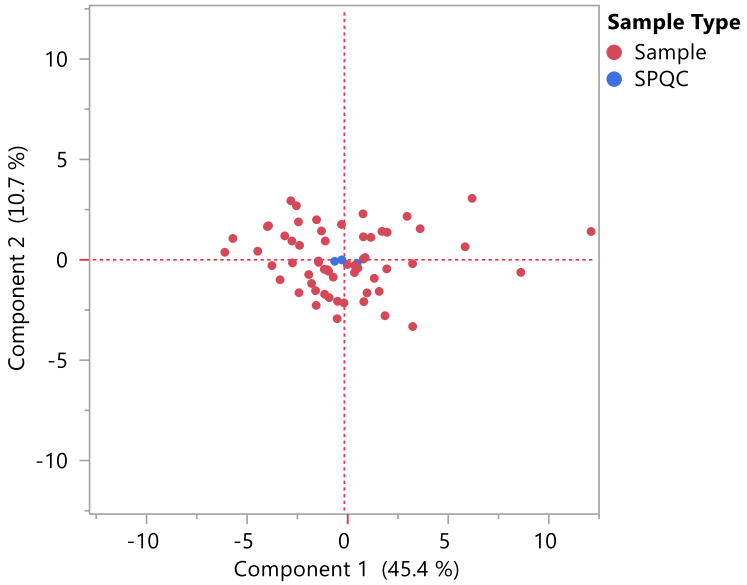
In order to assess general variability for the samples within each analyte class, to assess batch effect, and to look for sample outliers; a Principal Components Analysis (PCA) was performed for amino acids, biogenic amines, acylcarnitines, glycerophospholipids, and sphingolipids using JMP® Pro v13.0 software (SAS, Cary, NC) and the data from tables ***4745 UPLC p400 Data.xlsx*** and ***4745 FIA p400 Data.xlsx***. Analytes with measurable concentrations for more than half of the samples were included in the analysis. For analytes for which there were missing values, missing values were replaced with the Biocrates LOD value for that analyte and the restricted maximum likelihood (REML) method was used for correlation.

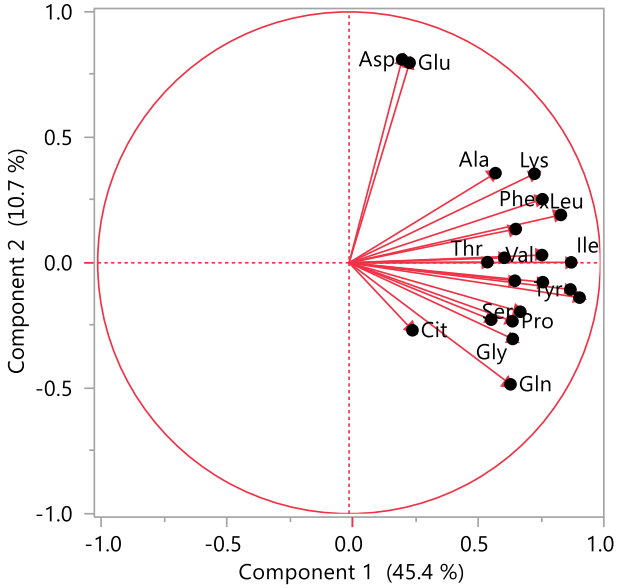
Figures below show PCA plots for each analyte class. Separate colors were assigned to samples (red) and 4745 SPQC (blue) in order to evaluate variability and to check for outliers.

The results for the pools are generally clustered together centrally within the samples which implies that there is greater biological than technical variability. Multivariate outlier screening using Mahalanobis distance in JMP Pro 13 revealed that there may be a few outlier samples when evaluating different classes of compounds independently (e.g. MD>15), however it is unclear whether there are any true outliers since those samples with MD>15 tend to be different when looking at the different classes of analytes. Few if any samples are outliers when looking at all analyte classes.

Principal Components Analysis of Amino Acids

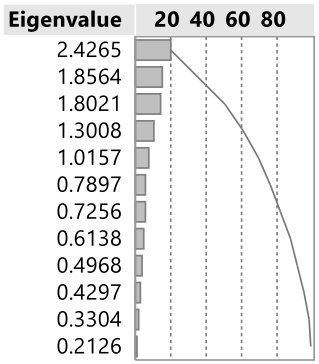


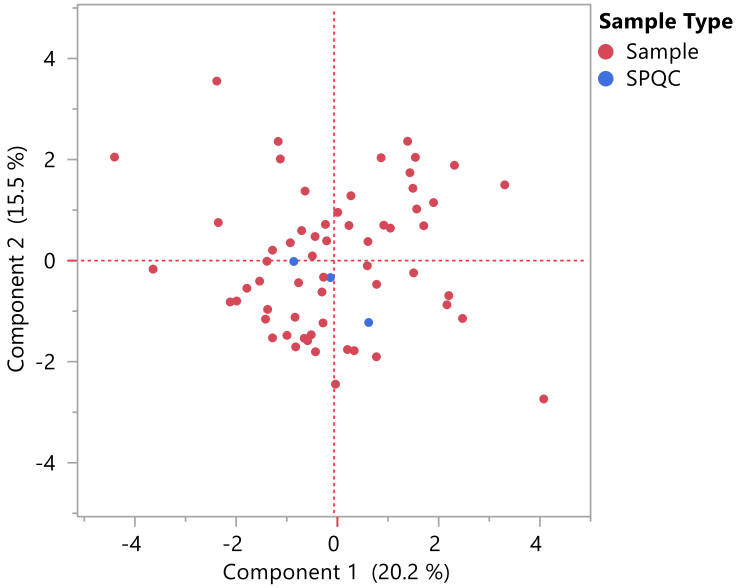


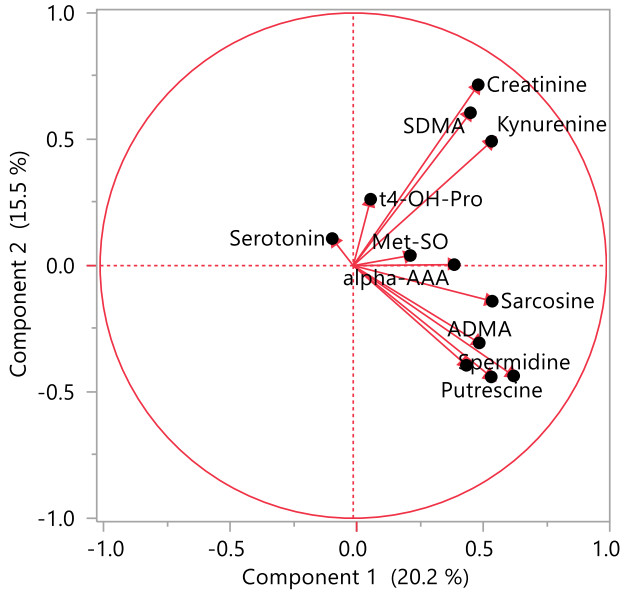


.

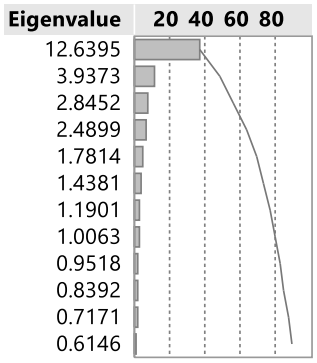
Principal Components Analysis of Biogenic Amines

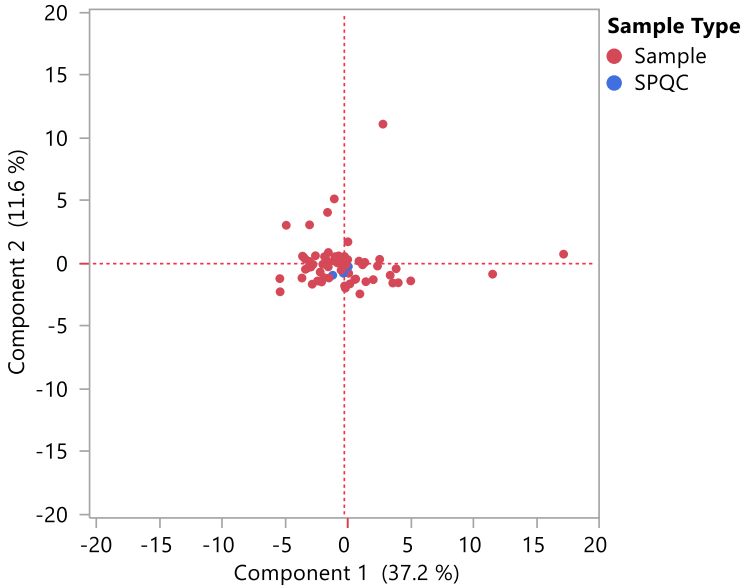


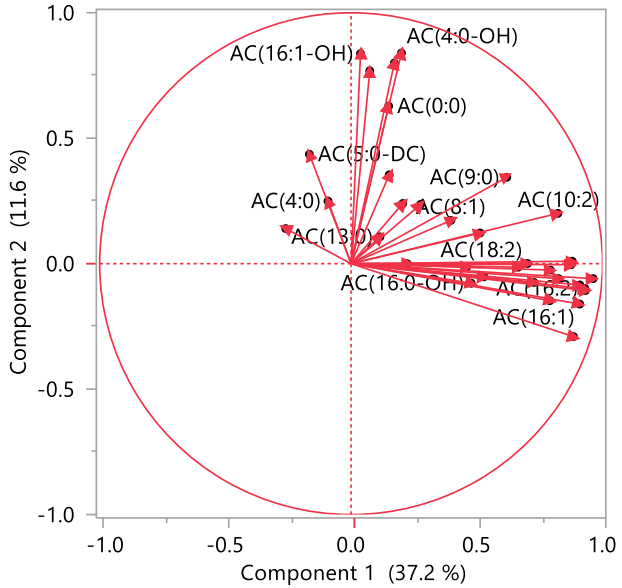




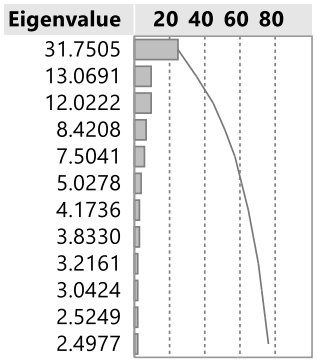
Principal Components Analysis of Acylcarnitines

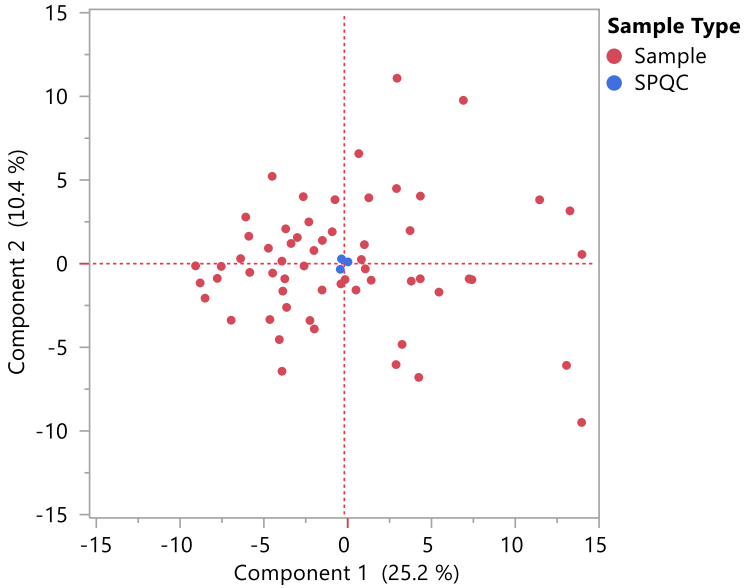


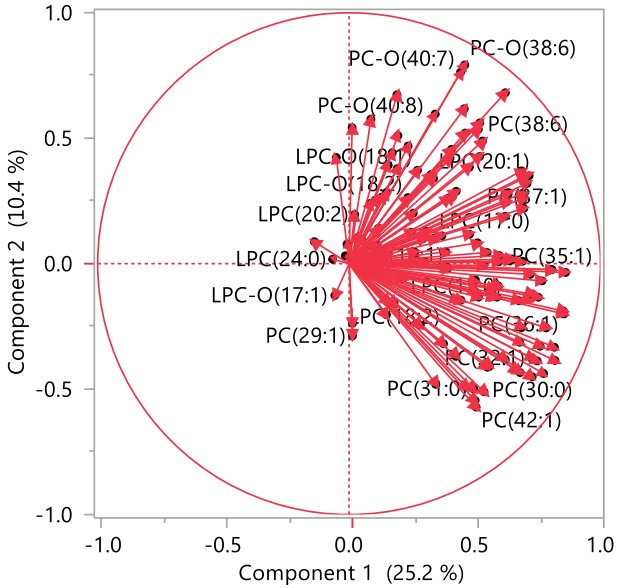




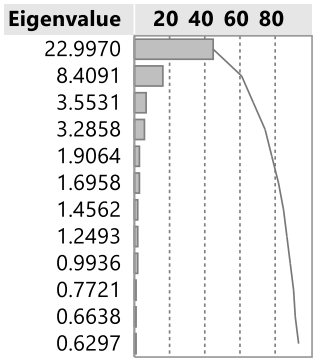
Principal Components Analysis of Glycerophospholipids

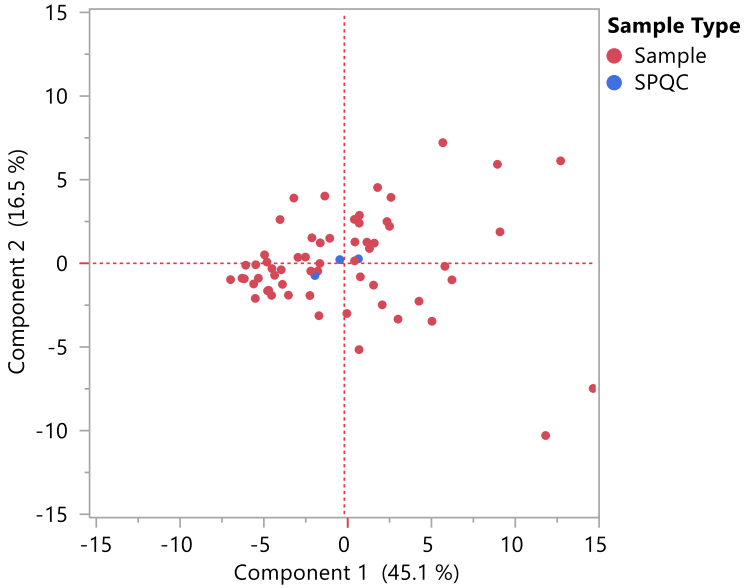


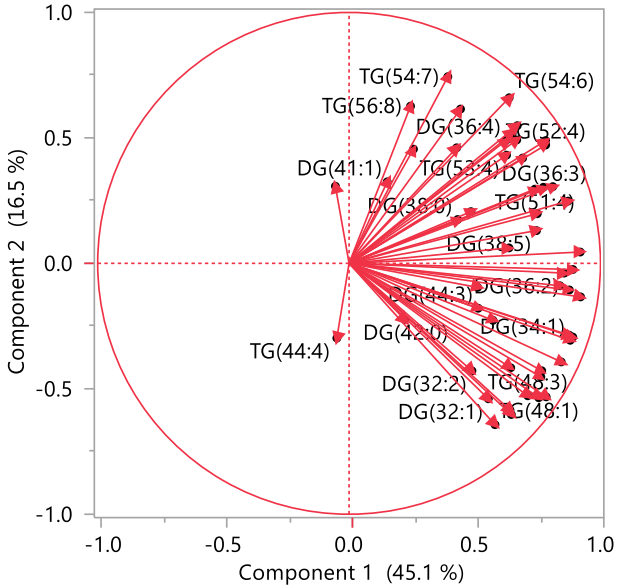




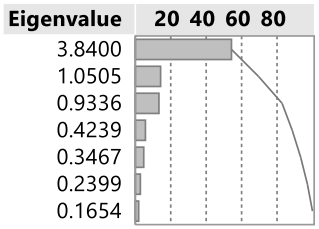
Principal Components Analysis of Glycerides

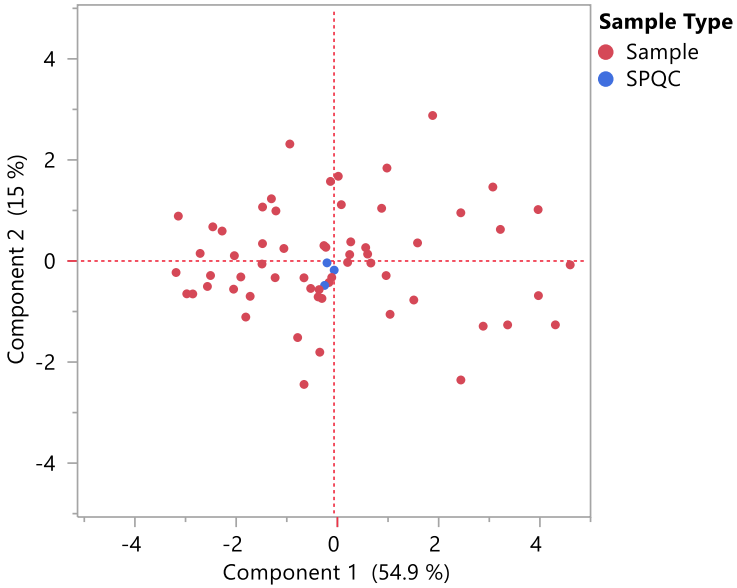


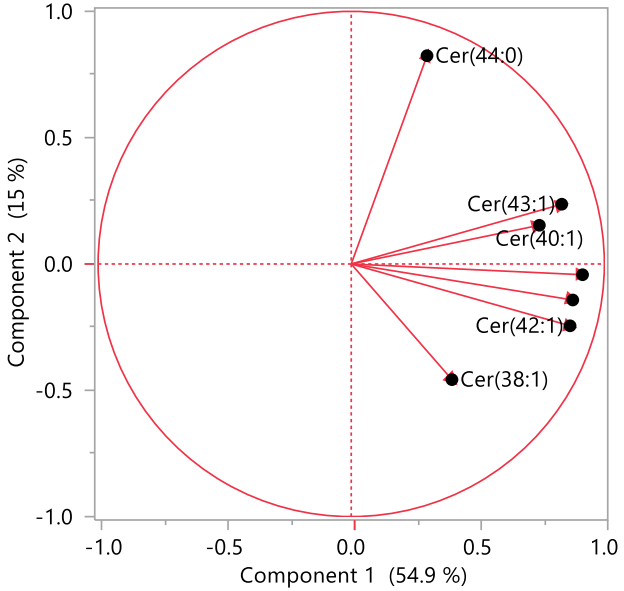




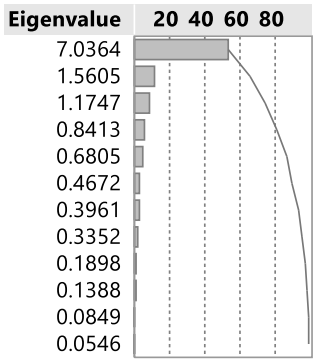
Principal Components Analysis of Ceramides

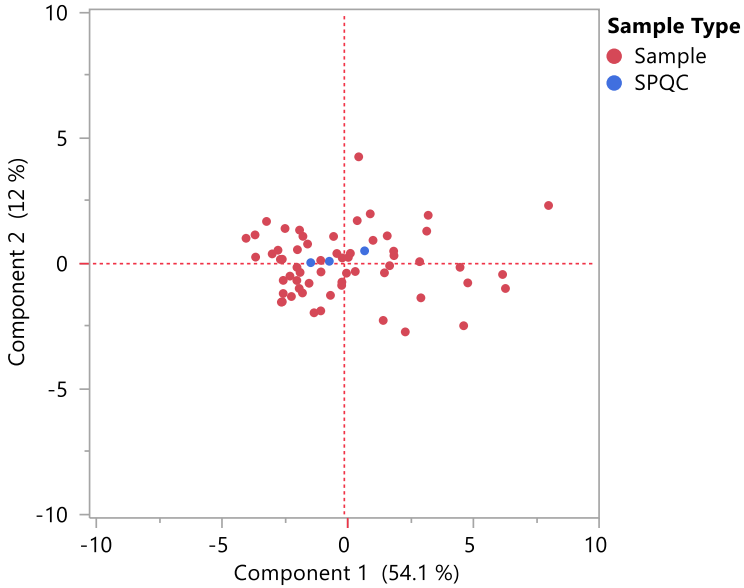


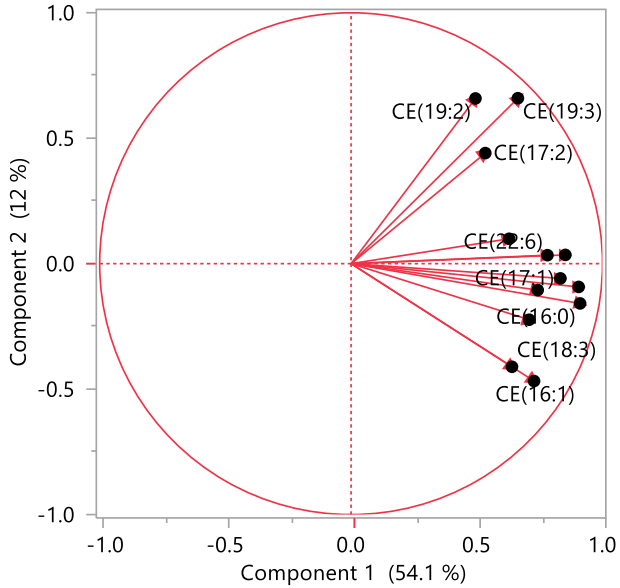




Principal Components Analysis of Cholesteryl Esters







Principal Components Analysis of Sphingolipids