Supplementary Information (SI) for Analyst.
This journal is © The Royal Society of Chemistry 2024

# Supplementary Information for

# Evaluating Protocols for Reproducible Targeted Metabolomics by NMR

Darcy Cochran,[1,2] Panteleimon G. Takis,[3,4*] James L. Alexander,[5,6,7] Benjamin H. Mullish,[5,6] Nick Powell,[5,6] Julian R. Marchesi[6] and Robert Powers[1,2*]

[1]*Department of Chemistry, University of Nebraska-Lincoln, Lincoln, Nebraska, 68588-0304, USA*

[2]*Nebraska Center for Integrated Biomolecular Communication, University of Nebraska-Lincoln, Lincoln, Nebraska, 68588-0304, USA*

[3]*Department of Chemistry, University of Ioannina, Ioannina GR 451 10, Greece*

[4]*Section of Bioanalytical Chemistry, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW7 2AZ, UK National Phenome Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, UK*

[5]*Departments of Gastroenterology and Hepatology, St Mary's Hospital, Imperial College Healthcare NHS Trust, South Wharf Road, Paddington London, W2 1NY, UK*

[6]*Division of Digestive Diseases, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, W2, 1NY, UK*

[7]*Department of Gastroenterology, St. Mark's Hospital and Academic Institute, Middlesex, UK*

*To whom correspondence should be addressed:

Panteleimon G. Takis
University of Ioannina
Department of Chemistry
Dourouti 451 10
Ioannina, Greece
Email: ptakis@uoi.gr
Phone: (+30) 2651008348

Robert Powers
University of Nebraska-Lincoln
Department of Chemistry
722 Hamilton Hall
Lincoln, NE 68588-0304
Email: rpowers3@unl.edu
Phone: (402) 472-3039
Fax: (402) 472-9402

**Keywords:** nuclear magnetic resonance, harmonization, sample preparation, methods, metabolomics

## EXPERIMENTAL DETAILS

### Study Design and Cohort Demographics

Plasma samples were acquired from an IBD cohort consisting of 69 patients and healthy volunteers recruited with approval from the research ethics committee (Ref. No. 21/WA/0105). Two identical ~1 mL aliquots of plasma were obtained from each of the 24 healthy age and sex matched controls, 21 patients with Crohn's disease, and 24 patients with ulcerative colitis for a total of 138 plasma samples. Demographic data was gathered at the time of sample collection and is provided in **Table S1**.

To evaluate the effect of sample preparation on metabolite profile detection, three unique protein removal procedures were applied to each individual plasma sample for comparison by 1D $^{1}$H NMR. Protein was removed from the plasma samples by either ultrafiltration, methanol precipitation, or the $T_2$ filtering of intact samples (**Figure S1**). Each sample preparation protocol was used as previously described by Gowda *et al.* (2014) in order to facilitate a direct comparison between these and other previously published results and the data obtained herein.[1-3] The plasma samples were thawed on ice and randomized before the preparation of the 219 NMR samples, consisting of 3 sample preparation protocols for each of the 69 unique plasma samples and the 4 pooled quality control (QC) samples. All sample preparation procedures were completed on the same day to reduce variation and unintended bias. 50 mM phosphate buffer in 100% $D_2O$ at pH 7.2 (uncorrected) with 100 μM of 3-(trimethylsilyl)propanoic-2,2,3,3-$D_4$ acid (TMSP) was prepared fresh and immediately before NMR sample preparation.[4]

### Preparation of Pooled QC Samples

A pooled QC sample was prepared for each of the three clinical groups, Crohn's disease, ulcerative colitis, and healthy controls. 55 µL was removed from each plasma sample from each clinical group and combined into a clean 2 mL centrifuge tube for a total volume ranging between ~1,155 to 1,320 µL for each of the three pooled QC samples. A fourth QC sample was prepared by combining 350 µL aliquots from each of the three individual QC samples. A total of 4 QC samples were then added to the original 69 clinical samples comprised of 21 Crohn's disease, 24 ulcerative colitis, and 24 healthy control samples for a combined total of 73 plasma samples.

**Preparation of NMR Samples Using Ultrafiltration**

Amicon Ultra centrifugal filters with a 3 kDa cutoff (UFC500396) were purchased from Sigma Aldrich (Saint Louis, MO) and washed 3x with 500 µL Nanopure water (Thermo-Fisher Scientific, Waltham, MA) at 14,000 × *g* for 20 minutes for a total centrifugation time of 60 minutes. 300 µL from each of the 73 plasma samples was transferred to an individual clean filter unit and centrifuged at 14,000 × *g* for 20 minutes and then the filtrate was dried overnight in a SpeedVac. Each dried sample was reconstituted in 300 µL Nanopure water and 300 µL of a fresh 50 mM phosphate buffer pH 7.2 (uncorrected) in 100% $D_2O$ with 100 µM TMSP-$D_4$ for a final TMSP concentration of 50 µM. Samples were vortexed and then 550 µL was transferred to a 7" Norell Standard Series 5 mm NMR tube (Morganton, NC, USA).

**Preparation of NMR Samples Using Methanol Precipitation**

A 300 µL aliquot from each of the 73 plasma samples was placed into an individual 2 mL centrifuge tube before adding 600 µL of ice-cold methanol to precipitate the protein. Samples were incubated at -20°C for 20 minutes before centrifugation at 14,000 × *g* for 30 minutes. The supernatant was

transferred to a fresh 2 mL centrifuge tube and then dried overnight in a SpeedVac. Each dried sample was reconstituted in 300 µL Nanopure water and 300 µL of a fresh 50 mM phosphate buffer pH 7.2 (uncorrected) in 100% $D_2O$ with 100 µM TMSP-$D_4$ for a final TMSP concentration of 50 µM. Samples were vortexed and then 550 µL was transferred to a 7” Norell Standard Series 5 mm NMR tube (Morganton, NC, USA).

**Preparation of NMR Sample Using an Intact Plasma Sample**

A 300 µL aliquot from each of the 73 plasma samples was placed into an individual 2 mL centrifuge tube. 300 µL of a fresh 50 mM phosphate buffer pH 7.2 (uncorrected) in 100% $D_2O$ with 100 µM TMSP-$D_4$ for a final TMSP concentration of 50 µM was added and then mixed by vortexing. 550 µL was transferred to a 7” Norell Standard Series 5 mm NMR tube (Morganton, NC, USA).

**NMR Data Acquisition**

All NMR experiments were performed at 298K on a Bruker Neo 600 MHz NMR spectrometer (Bruker BioSpin, Billerica, MA) equipped with a 5 mm TCI-F cryoprobe, temperature controlled SampleCase automated sample changer, ICON NMR and an automatic tune and match (ATM). All 1D $^{1}H$ NMR experiments were collected with 1D $^{1}H$ nuclear Overhauser effect spectroscopy (NOESY, noesygppr1d)[5] or a Carr-Purcell-Meiboom-Gill (CPMG, cpmgpr1d)[6, 7] pulse sequence, 65K data points, a spectral width of 17857 Hz, 64 scans, 16 dummy scans, and a 4s relaxation delay.

**Data Processing**

The 1D $^{1}H$ NMR spectra were zero-filled to 132K data points and Fourier transformed following an exponential apodization function of 1 Hz. Baseline and phase corrections were done via the Bruker automation software IconNMR 5.2.3.1 for TopSpin 4.1.3. When needed, a manual zero-order and first-order phase correction were applied, followed by baseline correction via fitting a third-degree polynomial function to regions of the spectrum lacking peaks. Both manual phase and baseline corrections were performed via TopSpin "apk" and “abs” functions, respectively.

Processed spectra were imported into Chenomx NMR Suite Professional Software Package (version 8.3; Chenomx Inc., Edmonton, Alberta, Canada) for metabolite quantification. Chenomx allows users to upload and process raw or pre-processed spectra, and then complete metabolite identification and quantification by comparing NMR peaks in the sample to their proprietary reference libraries as well as the libraries that have been developed in a collaborative effort with HMDB. Chenomx v8.3 offers three methods of metabolite identification that vary in degree of required manual human intervention. A fully automated, completely hands-free method of metabolite detection is possible by using the “batch-fit” method and choosing a library or metabolite list that Chenomx subsequently attempts to fit to the acquired NMR spectra of interest. A second Chenomx process is the “assisted-fit” method which allows users to upload a reference spectrum with metabolites of interest already fit to it that Chenomx then uses as a template for fitting the same metabolite profiles to other 1D $^{1}H$ NMR spectra in the data set. In this manner, the assisted-fit method allows for a user-directed semi-automated analysis. Finally, the third Chenomx method involves a fully manual approach to confirm the proper fitting of each reference 1D $^{1}H$ NMR spectra to each experimental NMR spectrum. Manual fitting was completed in combination with either the batch-fit or assisted-fit method used as a first step followed by manual adjustments as needed. A fully manual-fitting approach is highly time-consuming and is not always feasible

when analyzing large-scale studies consisting of hundreds or thousands of samples. Each of these three methods, batch-fit, assisted-fit, and manual-fit was applied to the dataset to evaluate their similarities and differences. The 27 metabolites selected for comparison and assessment are described in **Table S2**, which was based upon the panel of abundant metabolites previously described in the literature.[8] Glycerol and methanol were excluded from all further analysis since the compounds may be contaminants introduced by the sample preparation methodology leaving a total of 25 metabolites for comparison.

First, a blind batch-fit of the entire 600 MHz Chenomx library was applied to the complete dataset. A metabolite was excluded if it was not detected in 80% or more of the samples comprising a clinical group. An assisted-fit was conducted by manually fitting the clusters correlated to the selected set of 25 metabolites for a single reference 1D $^1$H NMR spectra for each clinical group. This manually fitted spectrum was then uploaded as a template starting point for Chenomx to complete the transformations and assignments automatically. A full manual-fitting of the reference 1D $^1$H NMR spectra for a select set of 25 metabolites to each of the experimental 1D $^1$H NMR spectra for each clinical group was completed for spectra acquired with the NOESY pulse sequence. The fit of each metabolite reference 1D $^1$H NMR spectra was individually evaluated and the Chenomx fit was manually adjusted to properly match chemical shifts and peak intensities.

The data analysis was also completed in parallel using the SMolESY platform with a semi-automated configuration.[9] Briefly, as previously shown[9] and validated in more than 8000 blood NMR spectra, the integral of a SMolESY signal of at least one $^1$H spin system from the 23 metabolites can be automatically assigned and integrated to provide the relative concentration of each plasma/serum metabolite. Consequently, we employed the same strategy via SMolESY-platform which obtains the assignment and quantification of the metabolites in a semi-automated

manner (for more details see User's guide semi-automated peak picking function of the software: https://github.com/pantakis/SMolESY_platform).

**Statistical Analysis**

Prior to statistical analysis, all data was range scaled to make a comparative analysis possible between Chenomx and SMolESY. The Chenomx manual-fit analysis of the 1D $^{1}$H NMR spectra acquired from the methanol-induced protein precipitation plasma samples was used as the standard data set for relative comparisons. Microsoft Excel and JMP 17.2.0 were used for data analysis. Group differences were statistically evaluated with one-way ANOVA testing followed by Tukey's posthoc test or FDR correction via the Benjamini-Hochberg method. Statistical significance was set at $p < 0.05$. Outlier testing was performed using Tukey's interquartile range (IQR) method and using 1.5 as the Q value for determining the range's upper and lower bounds.

## SUPPORTING REFERENCES


(1) Nagana Gowda, G. A.; Raftery, D. Quantitating metabolites in protein precipitated serum using NMR spectroscopy. *Anal Chem* **2014**, *86* (11), 5433-5440. DOI: 10.1021/ac5005103 From NLM Medline.

(2) Tiziani, S.; Emwas, A. H.; Lodi, A.; Ludwig, C.; Bunce, C. M.; Viant, M. R.; Gunther, U. L. Optimized metabolite extraction from blood serum for 1H nuclear magnetic resonance spectroscopy. *Anal Biochem* **2008**, *377* (1), 16-23. DOI: 10.1016/j.ab.2008.01.037 From NLM Medline.

(3) Madrid-Gambin, F.; Oller, S.; Marco, S.; Pozo, O. J.; Andres-Lacueva, C.; Llorach, R. Quantitative plasma profiling by (1)H NMR-based metabolomics: impact of sample treatment. *Front Mol Biosci* **2023**, *10*, 1125582. DOI: 10.3389/fmolb.2023.1125582 From NLM PubMed-not-MEDLINE.

(4) Kelly, A. E.; Ou, H. D.; Withers, W.; Dotsch, V. Low-Conductivity Buffers for High-Sensitivity NMR Measurements. *J. Am. Chem. Soc.* **2002**, *124* (40), 12013-12019.

(5) McKay, R. T. How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: An examination of the pulse sequence components and evolution. *Concepts in Magnetic Resonance Part A* **2011**, *38A* (5), 197-220. DOI: 10.1002/cmr.a.20223.

(6) Carr, H. Y.; Purcell, E. M. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Physical Review* **1954**, *94* (3), 630-638. DOI: 10.1103/PhysRev.94.630.

(7) Meiboom, S.; Gill, D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Review of Scientific Instruments* **1958**, *29* (8), 688-691. DOI: 10.1063/1.1716296.

(8) Takis, P. G.; Jimenez, B.; Al-Saffar, N. M. S.; Harvey, N.; Chekmeneva, E.; Misra, S.; Lewis, M. R. A Computationally Lightweight Algorithm for Deriving Reliable Metabolite Panel Measurements from 1D (1)H NMR. *Anal Chem* **2021**, *93* (12), 4995-5000. DOI: 10.1021/acs.analchem.1c00113 From NLM Medline.

(9) Takis, P. G.; Jimenez, B.; Sands, C. J.; Chekmeneva, E.; Lewis, M. R. SMolESY: an efficient and quantitative alternative to on-instrument macromolecular (1)H-NMR signal suppression. *Chem Sci* **2020**, *11* (23), 6000-6011. DOI: 10.1039/d0sc01421d From NLM PubMed-not-MEDLINE.

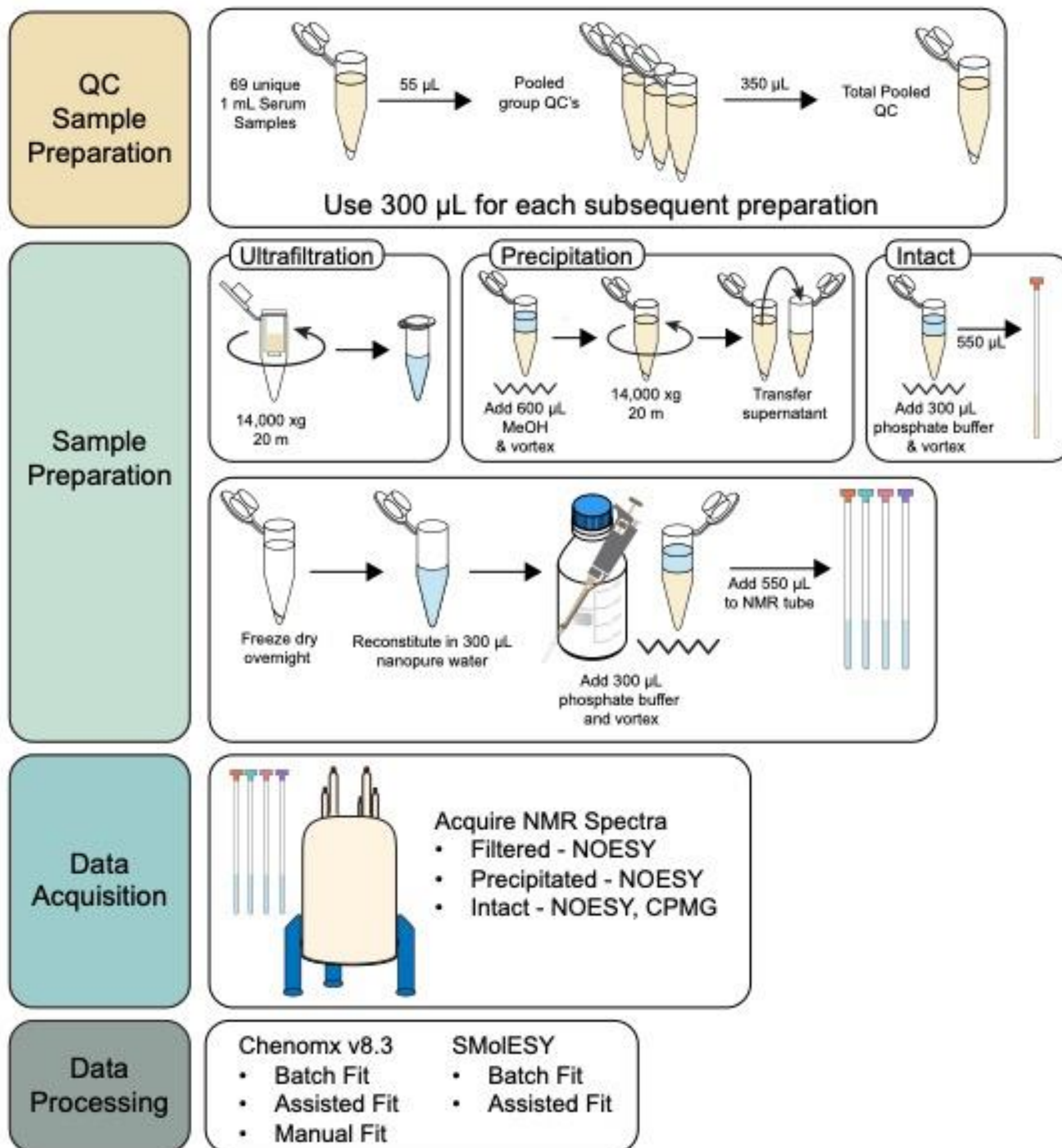


**Figure S1. Methodological Workflow**. Top to bottom stepwise diagram depicting the experimental workflow. 69 clinical samples and 4 QC samples were prepared with 3 unique sample preparation methods and analyzed by 2 different NMR experiments.

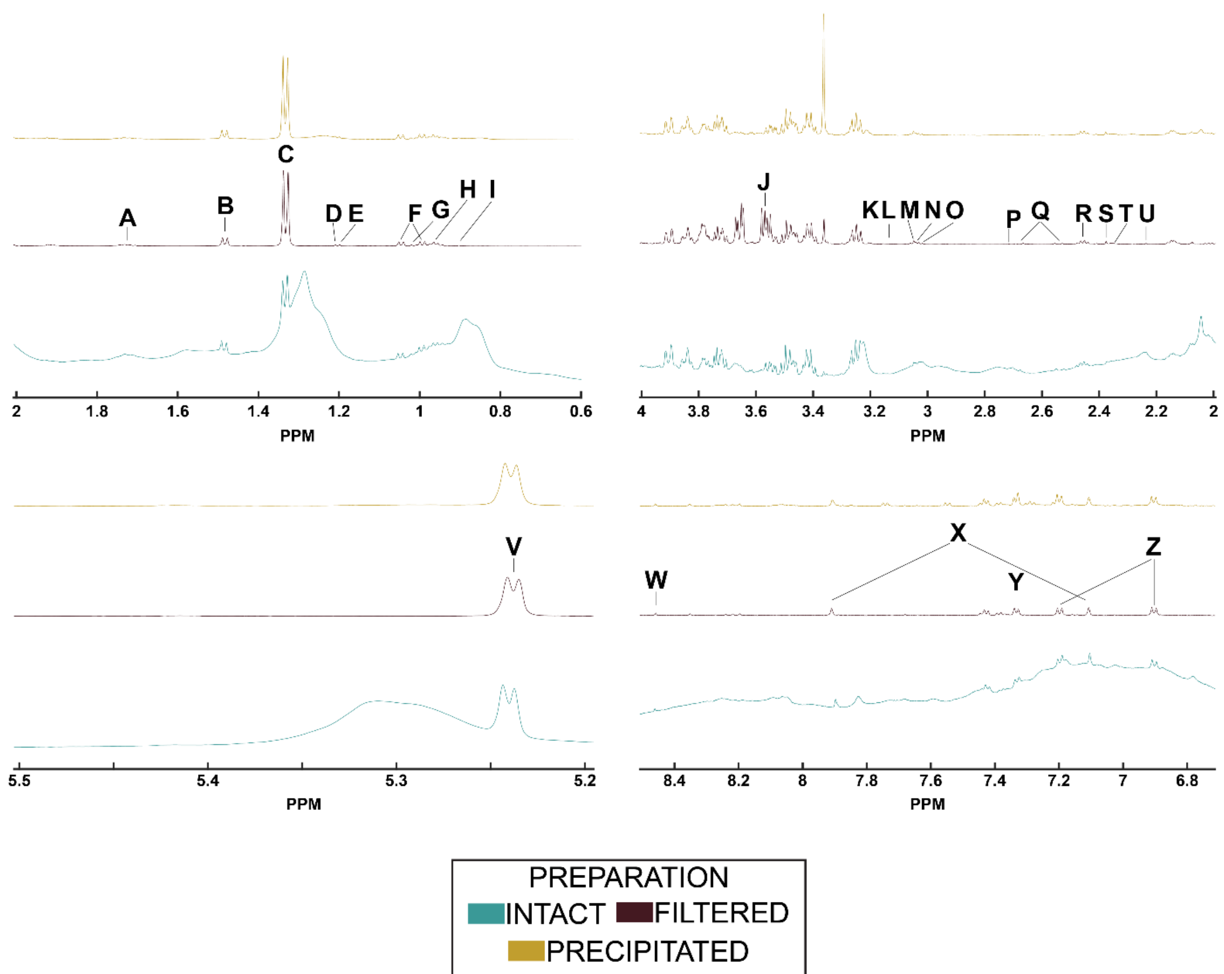


**Figure S2. Representative labeled1D $^{1}$H NMR Spectra**

1D $^{1}$H NMR spectra with labelled peaks corresponding to each metabolite of interest.

A – Acetate, B – Alanine, C – Lactate, D – 3-Hydroxybutyrate, E – Ethanol, F – Valine, G – Isoleucine, H – Leucine, I – 2-Hydroxybutyrate, G – Glycine, K – Choline, L – Dimethyl Sulfone, M – Creatinine, N – Creatine, O – Lysine, P – Dimethylamine, Q – Citrate, R – Glutamine, S – Pyruvate, T – Glutamate, U – Acetone, V – Glucose, W – Formate, X – Histidine, Y – Phenylalanine, Z - Tyrosine

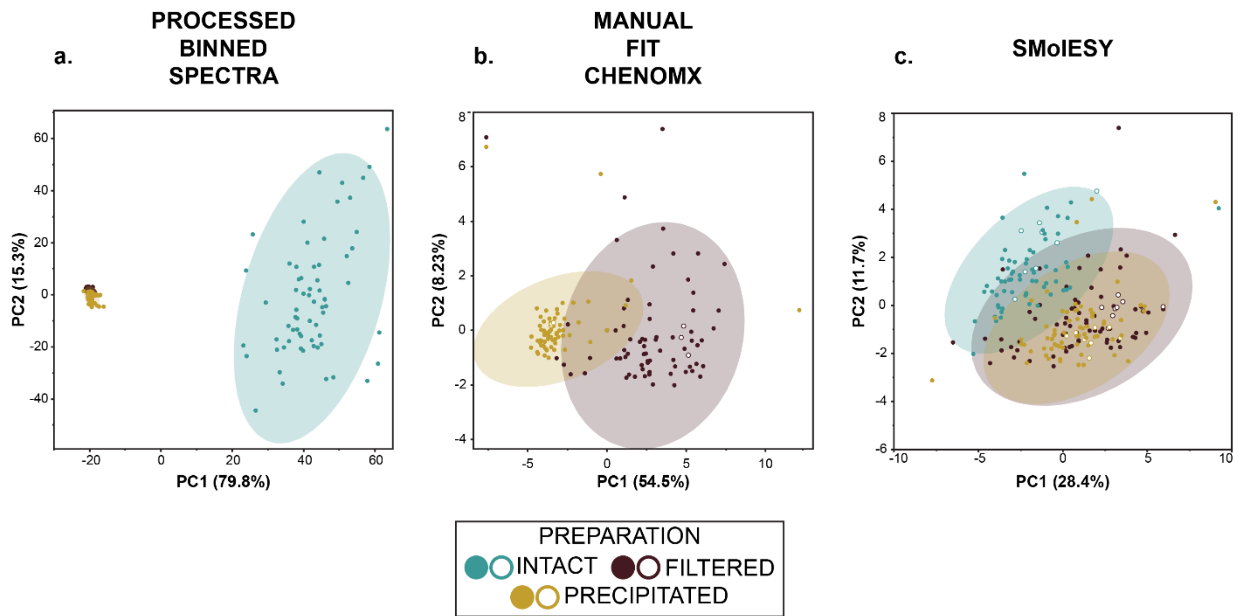


**Figure S3. PCA Group Separation.**

PCA scores plots demonstrating intact (teal) sample preparation group separation from ultrafiltered (burgundy) and methanol-induced protein precipitated (gold) groups using three data types **a.** binned spectral data, **b.** exported metabolite concentrations derived from manual Chenomx analysis and **c.** peak areas acquired from SMolESY analysis. Pooled quality control samples are open circles, clinical samples are filled circles. No QC samples are shown in panel **a** since the clusters are too compact.

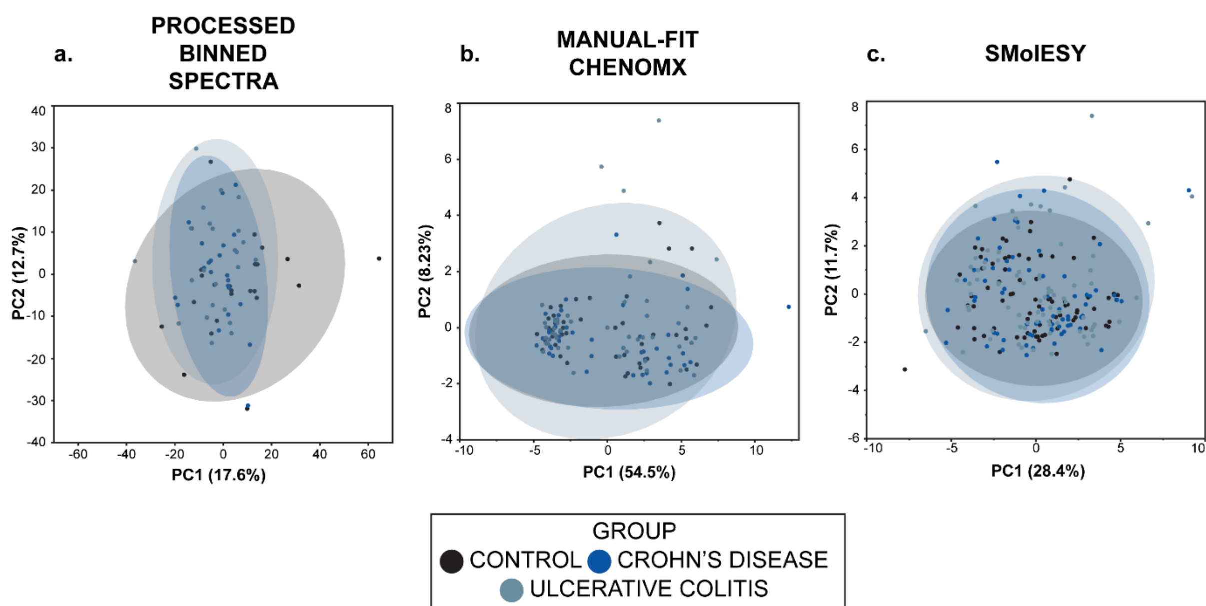


**Figure S4. PCA group separation by clinical grouping.**

PCA scores plots demonstrating the absence of group separation in a PCA model calculated with samples belonging to the control group (black), Crohn's disease (blue), or Ulcerative Colitis (blue-grey). PCA scores plot are obtained when the PCA models were generated with **a.** binned spectral data, **b.** exported metabolite concentrations derived from manual-fit Chenomx analysis or **c.** peak areas acquired from SMolESY analysis.

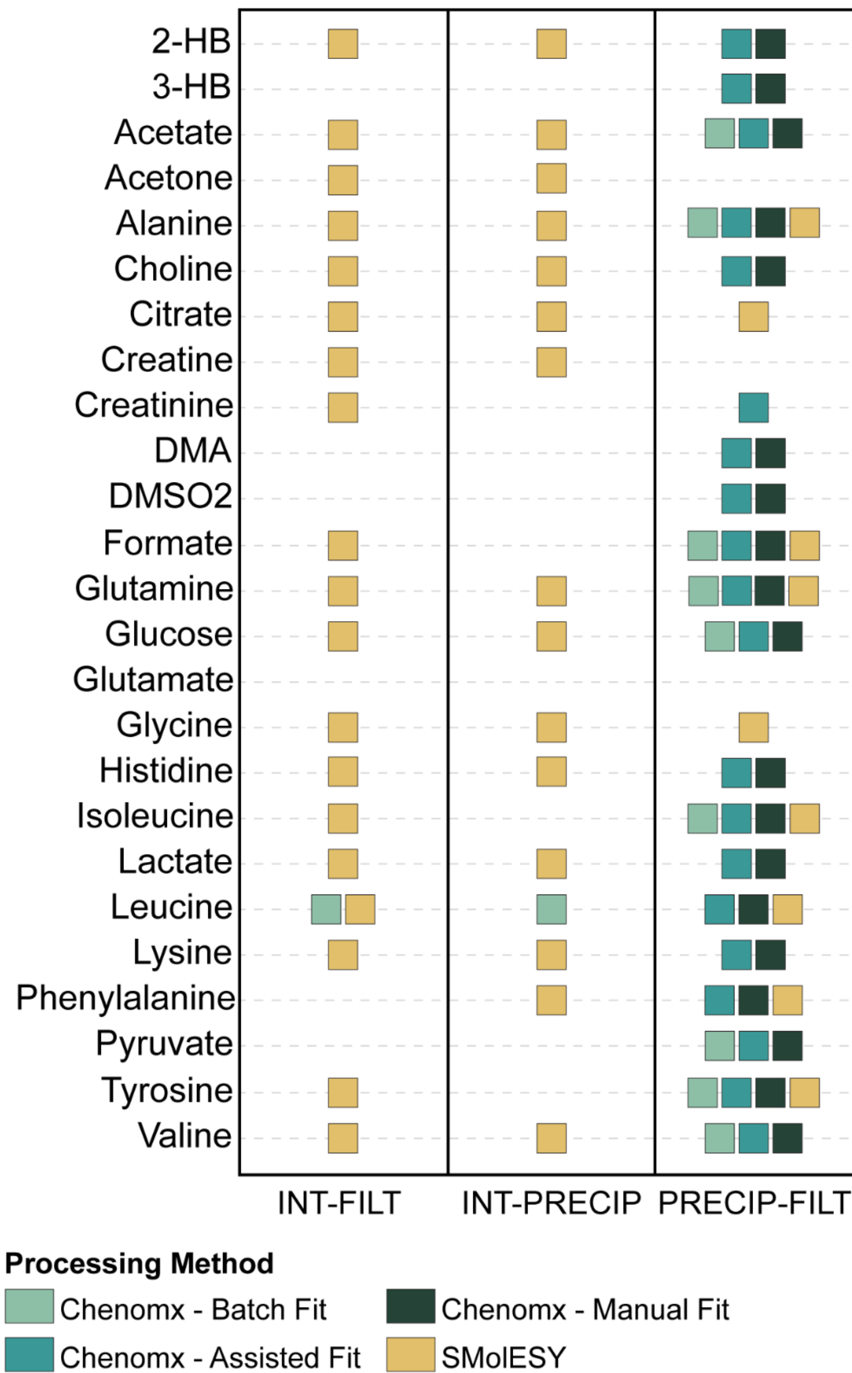


**Figure. S5** Pairwise Statistical Significance Colorized Table

Color-coded table lists metabolites in alphabetical order from top to bottom with statistical significance noted by presence or absence of colored square in the corresponding row. Presence of a square in a metabolite row denotes that it was found to be statistically different ($p < 0.05$) when comparing intact

(INT) samples to ultrafiltered (FILT) samples (INT-FILT, left), intact samples to methanol-induced protein precipitated (PRECIP) samples (INT-PRECIP, middle), or methanol-induced protein precipitated to ultrafiltered samples (PRECIP-FILT, right). The color of the square denotes the processing method the metabolite was identified as being statistically different: Chenomx–Batch fit (light blue), Chenomx–assisted fit (teal), Chenomx–Manual fit (dark green), SMolESY (gold). Abbreviations: 2-HB: 2-hydroxybutyrate, 3-HB: 3-hydroxybuytrate, DMSO2: dimethyl sulfone, DMA: dimethylamine.

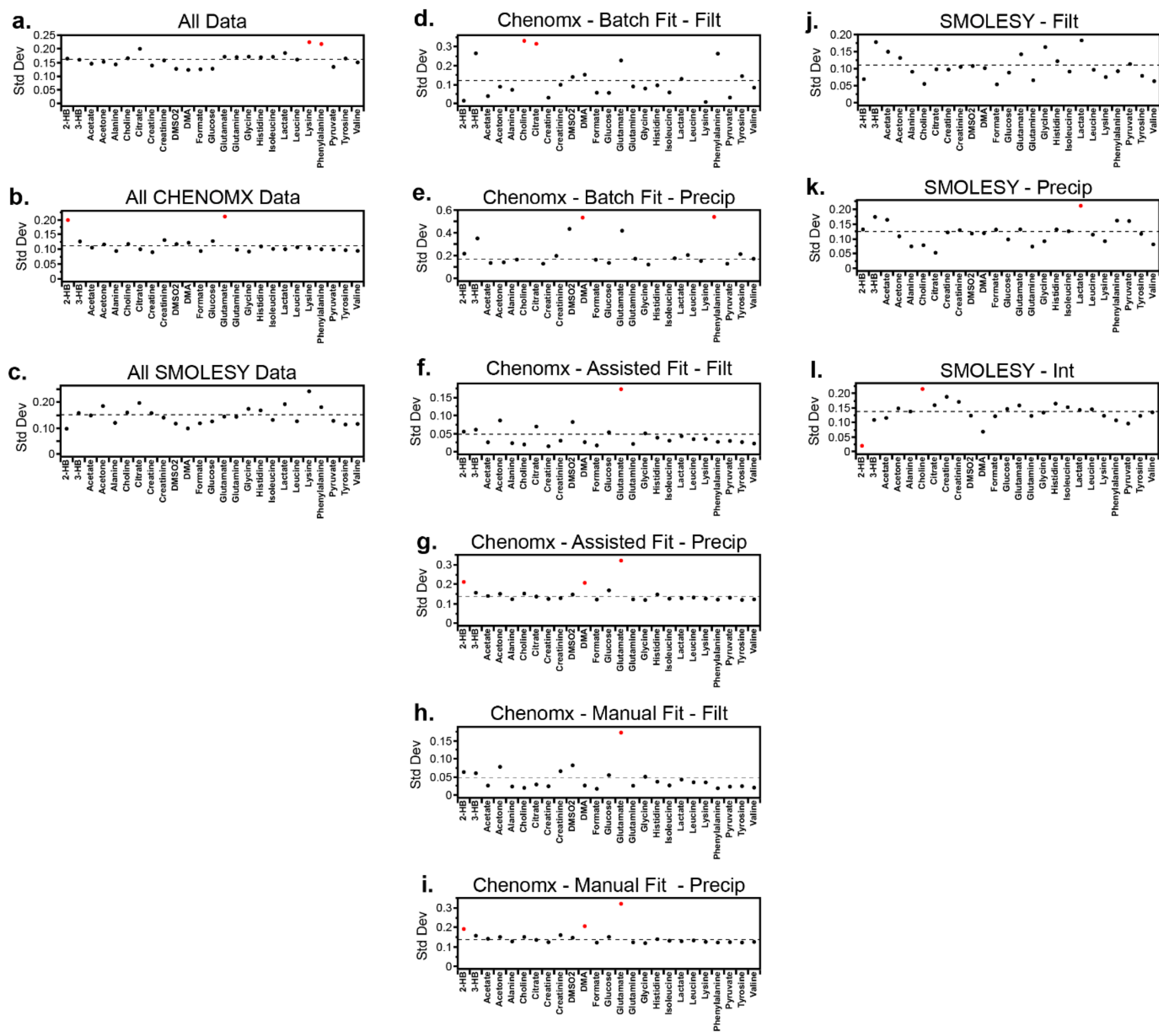


**Figure S6. Metabolite Variance by Preparation and Processing Method**

Scatterplots of standard deviations in metabolite concentrations as determined from each processing method. Outlier points are colored red. **a.** Total standard deviations from the combination of all data (Chenomx and SMolESY). **b.** Standard deviations from combined Chenomx assisted-fit and manual-fit data. **c.** Standard deviations from all SMolESY data. **d-i.** Standard deviations from individual Chenomx fit methods (**d-e**, batch-fit; **f-g**, assisted-fit; **h-I**, manual-fit). **j-l.** Standard deviations from SMolESY analysis of each sample preparation group (**j**, ultrafiltered; **k**, methanol-induced protein precipitated; **l**, intact).