analytical chemistry

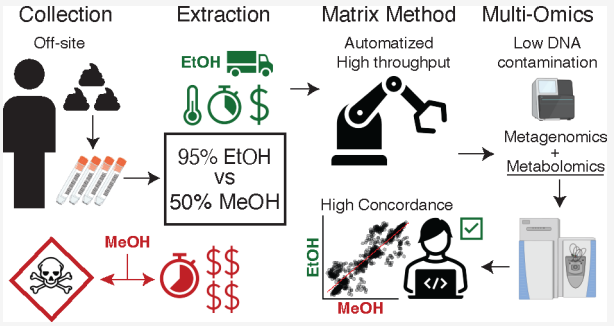
pubs.acs.org/ac
Article

Human Untargeted Metabolomics in High-Throughput Gut Microbiome Research: Ethanol vs Methanol

Simone Zuffa,* Vincent Charron-Lamoureux, Caitriona Brennan, Madison Ambre, Rob Knight, and Pieter C. Dorrestein

Cite This: *Anal. Chem.* 2025, 97, 4945−4953
Read Online

ABSTRACT: Untargeted metabolomics is frequently performed on human fecal samples in conjunction with sequencing to unravel the gut microbiome functionality. As sample collection efforts are rapidly expanding, with individuals often collecting specimens at home, metabolomics experiments should adapt to accommodate the safety and needs of bulk off-site collections and improve high throughput. Here, we show that a 95% ethanol, safe to be shipped and handled, extraction part of the Matrix Method pipeline recovers comparable amounts of metabolites as a validated 50% methanol extraction, preserving metabolic profile differences between investigated subjects. Additionally, we show that the fecal metabolome remains relatively stable when stored in 95% ethanol for up to 1 week at room temperature. Finally, we suggest a metabolomics data analysis workflow based on robust centered log ratio transformation, which removes the variance introduced by possible different sample weights and concentrations, allowing for reliable and integration-ready untargeted metabolomics experiments in gut microbiome research.



INTRODUCTION

Humans are colonized at birth by microorganisms, forming complex communities known as microbiota, which evolve through the course of the host life, shaping and influencing its physiology and metabolism.[1] Among others, the gut microbiota has been shown to actively influence neurodevelopment before[2,3] and after *partum*[4−6] and to induce distal tumors via carcinogenic metabolism.[7] For these reasons, elucidating the functionality of these microbial communities has become paramount, and mass spectrometry-based untargeted metabolomics has established itself as the to-go tool for these types of investigations.[8]

Multiomics large-scale longitudinal or cross-sectional microbiome studies represent a challenge both for sequencing and metabolomics as samples are usually collected in different settings and cannot always be immediately snap-frozen in liquid nitrogen and stored at −80 °C, which is considered the gold standard. Additionally, when designing large-scale studies, the safety of subjects and shipping costs, which dramatically increase if refrigeration is involved, should be taken into consideration. The microbiome field tackled these problems by showcasing that fecal microbiome collection and storage in 95% ethanol (EtOH) at room temperature stabilizes the microbial communities up to 8 weeks.[9] Most importantly, EtOH is also safer to handle when compared to other alcohols, such as methanol (MeOH), which is extremely toxic and requires special equipment to be properly handled. Building on this, we recently introduced the Matrix Method,[10] which employs a high-throughput pipeline that leverages sample collection in single barcoded Matrix tubes containing 95% EtOH and automatized robots. This method not only reduces costs, time, and well-to-well DNA contamination but also enables the extraction of metabolites from the same biological sample, offering an all-in-one solution streamlined multiomics analyses. EtOH extraction from fecal metabolomics studies has been tested before,[11−14] but an in-depth downstream data analysis comparison with another common extraction method (50% MeOH),[15−18] used for the discovery of novel bile acids conjugates[19] and the recently introduced reverse metabolomics approach,[20] was missing.

Here, we showcase that the 95% EtOH metabolomics extraction part of the Matrix Method for bulk microbiome analyses yields equivalent results of a common, but more laborious and time-consuming, 50% MeOH extraction for human fecal samples. We also show that 95% EtOH can stabilize the fecal metabolome at room temperature for up to 1 week, allowing for safe off-site collection and possibly reduced

Received: September 23, 2024
Revised: January 7, 2025
Accepted: February 18, 2025
Published: February 27, 2025

https://doi.org/10.1021/acs.analchem.4c05142
Anal. Chem. 2025, 97, 4945−4953

shipping costs. Finally, we highlight how a robust center log ratio (RCLR) transformation data analysis workflow for untargeted fecal metabolomics data overcomes problems of uneven sampling collection and allows for direct integration with microbiome data.

EXPERIMENTAL SECTION

Sample Collection and Extraction. Human fecal samples were collected and immediately stored at −80 °C from healthy volunteers under approved protocols from the University of California San Diego (IRB#141853) with informed consent. Three fecal samples from three different adult subjects (one male and two females) were randomly selected and aliquoted for untargeted metabolomics analysis. Multiple aliquots of different weights (10, 20, and 30 mg) were generated in triplicates from the three different fecal samples. Aliquots were transferred in either 95% (v/v) ethanol (EtOH) or 50% (v/v) methanol (MeOH). Samples were immediately extracted, except for a subset of samples that were left in 95% EtOH for either 24 h or 1 week at room temperature to check the fecal metabolome stability. Samples to which 400 μL of 95% EtOH was added were extracted via the Matrix Method pipeline,[10] simply consisting in shaking samples at 1200 rpm for 2 min in a SpexMiniG plate shaker (SPEX SamplePrep part #1600, NJ, USA), followed by a 5 min centrifugation step at 2700*g*. The supernatant (400 μL) was then collected and stored at −80 °C for downstream analysis. Samples to which 800 μL of MeOH was added underwent a validated extraction protocol,[20] involving homogenization with a 5 mm stainless-steel bead in a TissueLyser II (QIAGEN) for 5 min at 25 Hz, incubation at 4 °C for 30 min, and centrifugation at 21,130*g* for 10 min. The supernatant (400 μL) was then collected and dried overnight using a CentriVap instrument. Samples were then stored at −80 °C until resuspension. All supernatants, from EtOH and MeOH extractions, were dried overnight using a CentriVap instrument and then resuspended in 200 μL of 50% MeOH containing 1 μM of sulfamethazine as the internal standard. A pooled sample (QCpool) was then generated by collecting and mixing 10 μL from each biological sample and aliquoting 200 μL. Blank samples, consisting only of extraction solution, were also prepared. Finally, the samples were incubated for 1 h at −20 °C, centrifuged at 21,130*g* for 10 min, and transferred in 2 mL glass vial (Thermo Scientific) for ultrahigh-performance liquid chromatography tandem mass spectrometry (UHPLC-MS/MS) analysis.

UHPLC-MS/MS Experiment. Samples were randomized and analyzed using an untargeted metabolomics analysis platform comprising a Vanquish UHPLC system coupled to a Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific). The chromatography system consisted of a Phenomenex C18 column (1.7 μm particle size, 2.1 mm × 50 mm) and a mobile phase of solvent A (water + 0.1% formic acid) and solvent B (acetonitrile + 0.1% formic acid). Injections of 5 μL of samples, with a flow rate of 0.5 mL/min, followed this gradient: 0−1 min, 5% B; 1−7 min, 5−99% B; 7−8 min, 99% B; 8−8.5 min, 99−5% B; 8.5−10 min, 5% B. MS/MS data were acquired in the data-dependent acquisition mode using positive electrospray ionization (ESI+). Briefly, ESI parameters were set as follows: 53 L/min sheath gas flow, 14 L/min aux gas flow rate, 3 L/min sweep gas flow, 3.5 kV spray voltage, 269 °C intel capillary, and aux gas heater set to 438 °C. MS scan range was set to 100−1500 m/z with a resolution at m/z 200 set to 35,000 with 1 microscans. Automatic gain control (AGC) was set to 5E4 with a maximum injection time of 50 ms. Up to 5 MS/MS (TopN = 5) spectra per MS1 were collected with a resolution at m/z 200 set to 17,500 with 1 microscans. Injection time was 50 ms with an AGC target of 5E4. The isolation window was set to 2.0 m/z. Normalized collision energy was set to a stepwise increase of 20, 30, and 40 eV with an apex trigger set to 2−15 s and a dynamic exclusion of 10 s.

UHPLC-MS/MS Data Processing. Obtained raw files were converted into .mzML open-access format using ProteoWizard MSConvert[21] and deposited on GNPS/MassIVE under the accession number MSV000095260. Feature detection and extraction was performed via MZmine 3.9 via batch processing.[22] The .xml file used for batch processing can be found on the associated GitHub page. Briefly, data were imported using MS1and MS2 detector via factor of lowest signal with noise factors set to 3 and 1.1, respectively. Sequentially, mass detection was performed, and only ions were acquired between 0.5 and 8 min, with MS1 and MS2 noise levels set to 5E4 and 1E3, respectively. Chromatogram builder parameters were set at five minimum consecutive scans, 1E5 minimum absolute height, and 10 ppm for m/z tolerance. Smoothing was applied before the local minimum resolver, which had the following parameters: chromatographic threshold of 85%, minimum search range retention time of 0.2 min, and minimum ratio of peak top/edge of 1.7. Then, the 13C isotope filter and isotope finder were applied. Features were aligned using join aligner with weight for m/z set to 3 and retention time tolerance set to 0.2 min. Features not detected in at least three samples were removed before performing peak finder. Ion identity networking and metaCorrelate were performed before exporting the final feature table. The GNPS and SIRIUS export functions were used to generate the feature table containing the peak areas and the .mgf files necessary for downstream analyses. Feature-based molecular networking was performed in GNPS2 (https://gnps2.org/status?task=40d2affb3df544d4a2bbf6841a62d45a#),[23] and it was used to annotate metabolic features via MS/MS spectral matching to the GNPS library, which represent a level 2 annotation according the Metabolomics Standard Initiative. Annotations were obtained via parent ion mass matching with a 0.02 tolerance, a minimum of five matching fragment ions, and a cosine score similarity > 0.7. A list of all annotated molecular features can be found in the "Library Results" tab of the FBMN job page. Molecular classes of ions with m/z < 800 were predicted using CANOPUS Natural Product Classifier (NPC) via SIRIUS 5.8.[24]

Data Analysis. Feature Table was imported in R 4.2.2 (R Foundation for Statistical Computing, Vienna, Austria) for downstream data analysis. Total extracted peak area per sample was calculated and correlated with the sample run order to identify possible acquisition problems during the run. The internal standard (IS) peak present in each sample (Sulfamethazine [M + H], m/z 279.0908 and retention time 2.26 min) was also extracted and correlated to the sample run order. The coefficients of variance (CVs) of six different standards (amitriptyline, sulfadimethoxine, sulfamethazine, sulfamethizole, sulfachloropyridazine, and coumarin 314) present in the QCmix sample, which was run every 10 biological samples, were inspected to evaluate the run quality. The CVs of each acquired feature were also calculated using the QCpool,[25] which was also run for every 10 biological samples. The CV was calculated by dividing the mean of the

https://doi.org/10.1021/acs.analchem.4c05142
Anal. Chem. 2025, 97, 4945−4953

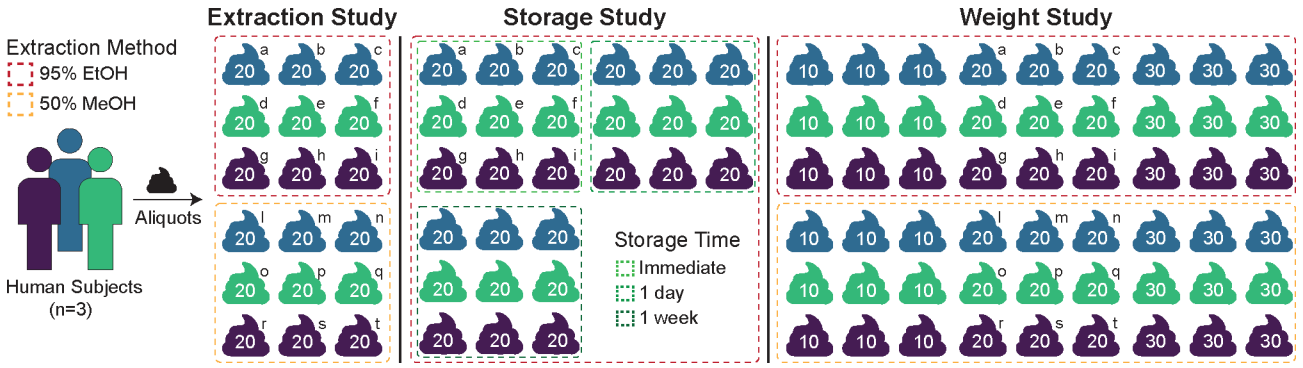


Figure 1. Study design. Fecal samples were collected from three different human subjects and aliquoted to investigate two different extraction pipelines: a 95% ethanol (EtOH) extraction part of the Matrix Method and a 50% methanol (MeOH) extraction protocol. Three different analyses were conducted. (1) In the extraction study, 20 mg of aliquoted fecal samples were immediately processed after thawing to investigate metabolites recovery differences between 50% MeOH and 95% EtOH pipelines. (2) In the storage study, 20 mg of aliquoted fecal samples were immediately processed or left at room temperature for 1 day and 1 week in 95% EtOH before undergoing the Matrix Method to investigate fecal metabolome stability in 95% EtOH. (3) In the weight study, 10, 20, and 30 mg of aliquoted fecal samples were extracted via both 95% EtOH and 50% MeOH pipelines to investigate a weight-bias free data analysis workflow. Colors indicate subject origin. Numbers (10, 20, 30) indicate weight in mg of the aliquoted fecal sample. The immediately processed 20 mg 95% EtOH samples (a–i) have been used in the extraction, storage, and weight study, while the immediately processed 20 mg of 50% MeOH samples (l–t) have been used in the extraction and weight studies.

extracted peak areas by the standard deviation. Feature table was cleaned via blank subtraction. Features only detected in the blank and QCmix samples or those that mean peak areas were not at least 10 times the ones observed in the QCpool were discarded. The package 'homologueDiscoverer v 0.0.0.9' was used to remove detected PEGs (polyethylene glycol) contaminants.[26] Features with near zero-variance were removed using 'caret v 6.0'.[27] The package 'mixOmics v 6.22' was used for multivariate analysis.[28] Principal component analysis (PCA) and partial least-squares discriminant analysis (PLS-DA) were performed after RCLR transformation via 'vegan v 2.6'.[29] In the PCA models, PERMANOVA was used to evaluate group centroid separation, while PERMDIPS2 was used to evaluate homogeneity of variance between groups. PLS-DA models' performances were evaluated using leave-one-out (loo) cross-validation. Variable importance (VIP) scores were calculated per feature and features with VIPs > 1 were considered significant. The package 'UpSetR v 1.4' was used to generate the upset plots.[30] High density region (HDR) plots were generated using 'ggdensity v 1.0'. Log2 fold changes (Log2FC) were calculated by taking the log2 of divided means of the relative abundance of the peak areas of the group of interest. When the mean was 0, a pseudocount of 1×10^{-9} was added. Linear mixed effect models were obtained using 'lmerTest v 3.1' using subject id as the random effect. The packages 'tidyverse v 2.0' and 'ggpubr v 0.6' were used for data manipulation and visualization. Code used for the analysis and to generate the figures of the manuscript is available on GitHub (https://github.com/simonezuffa/Manuscript_Matrix_Metabolomics).

RESULTS AND DISCUSSION

Fecal samples from 3 different human subjects were aliquoted to generate 72 replicates (Figure 1). Triplicates of different weights, 10, 20, and 30 mg, were generated and extracted via two different pipelines. The first one consisted of a 50% MeOH fecal extraction protocol, previously described,[20] while the second one involved a 95% EtOH extraction part of the recently introduced Matrix Method pipeline.[10] This consists of the collection of fecal samples in Matrix tubes containing 400 μL of 95% EtOH, followed by 2 min shaking at 1200 rpm and 5 min centrifugation at 2700*g*. Importantly, this automatized high-throughput pipeline allows for the simultaneous extraction of DNA, allowing multiomics analyses in microbiome research. Additionally, two batches of triplicates collected in 95% EtOH were left at room temperature for 1 day and 1 week, respectively, to replicate possible collection scenarios of storing, shipping, and assessing fecal metabolome stability in 95% EtOH.

95% EtOH Extraction Part of the Matrix Method Recapitulates a 50% MeOH Fecal Extraction Protocol. Unsupervised dimensionality reduction via PCA of 20 mg of triplicates showed clear clustering of subject fecal metabolic profiles using both 95% EtOH and 50% MeOH extractions (Figure 2A). PERMANOVA identified subject id as the highest source of variance in the data ($R^2 = 0.47$, $F = 10.98$, $p < 0.001$), followed by extraction method ($R^2 = 0.14$, $F = 6.68$, $p < 0.001$). Out of a total of 4616 unique metabolic features, 75% (3463) were recovered by both extraction methods (Figure 2B). These included 94% (235) of the total annotated features via the GNPS library. Interestingly, the Matrix Method (95% EtOH) appeared to capture a higher number of additional features (849) compared to that of the MeOH extraction (304). These extraction-specific features were classified as small peptides and fatty acid conjugates for MeOH and oligopeptides, glycerolipids, and fatty amides for EtOH according to CANOPUS NPC superclass predictions (Supplementary Table 1). Pairwise PCA and PLS-DA models were generated for each extraction method to identify the metabolic features responsible for subject discrimination. These features were then selected to compare both pipelines. All PCA and PLS-DA score plots displayed clear clustering by subject (Supplementary Figure 1). All PLS-DA models obtained a classification error rate of 0, indicating a perfect discriminating performance. Extracted VIP scores from the PLS-DA models displayed significant correlation between the two extraction pipelines (Supplementary Figure 2), with the highest concordance as observed in the HDR plots (Figure 2C). On average, the PLS-DA models identified 1764 features with VIPs > 1 involved in subject discrimination. Notably, when considering exclusively the top 100 features obtained in the

https://doi.org/10.1021/acs.analchem.4c05142
Anal. Chem. 2025, 97, 4945−4953

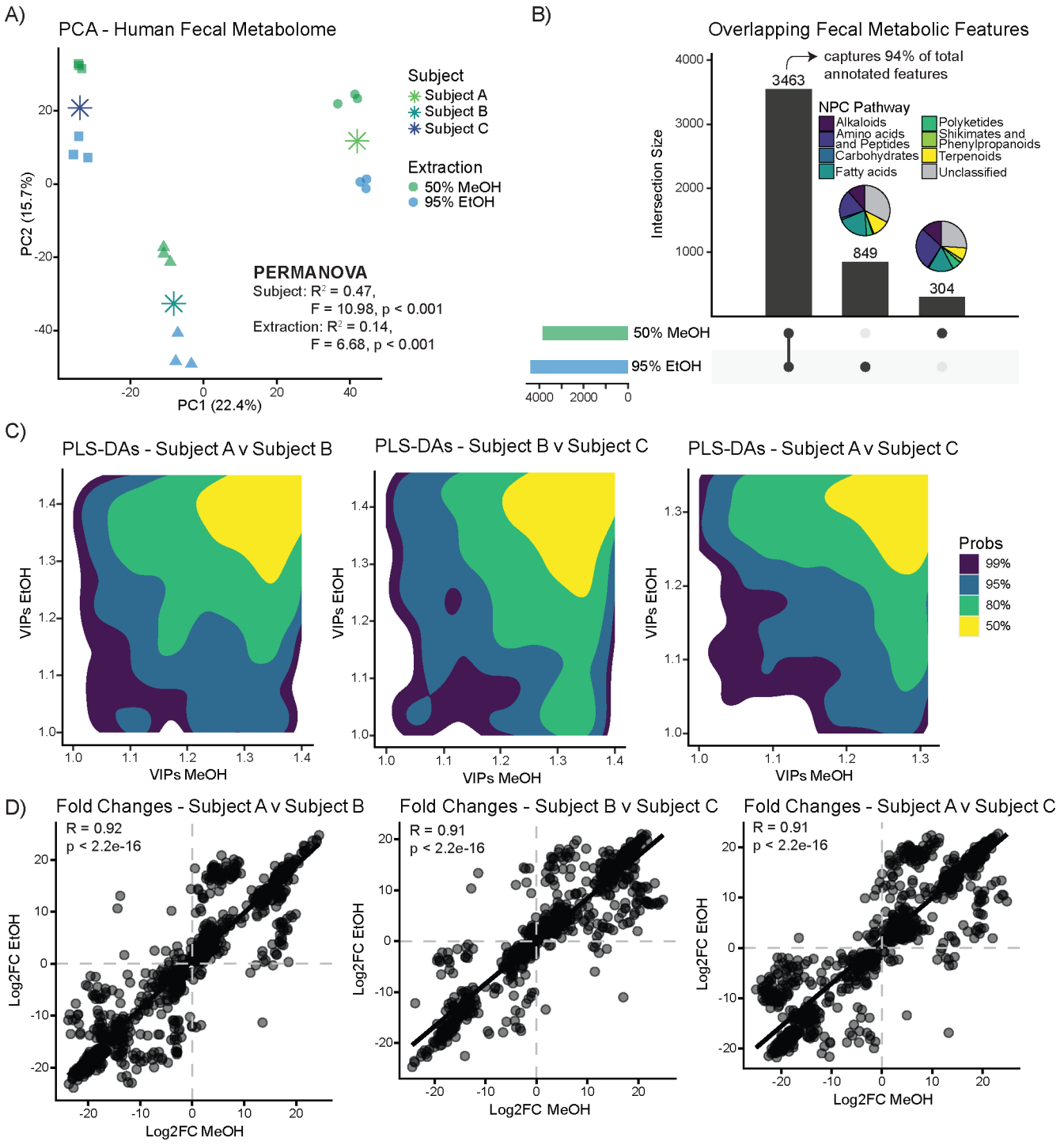


Figure 2. Data analysis comparison between 95% EtOH and 50% MeOH extraction pipelines. (A) PCA on RCLR-transformed peak areas shows clear clustering based on subject (PERMANOVA: subject R^2 = 0.47, p < 0.001 and extraction R^2 = 0.14, p < 0.001). (B) Top-right-hand side displays that more than 75% of the obtained features were recovered by both extraction pipelines, encompassing 94% of the total annotated features. Additionally, EtOH appeared to recover a higher number of oligopeptides, glycerolipids, and fatty amides, whereas MeOH recovered more small peptides and fatty acid conjugates (Supplementary Table 1). (C) HDR plots of feature VIP scores obtained via subject pairwise PLS-DA models stratified for the extraction method display a high degree of concordance. (D) Scatter plots of the pairwise Log2FC obtained via pairwise subject comparison in both extraction methods. Linear regression shows significant correlation (R > 0.9, p < 2.2 × 10^{-16}) between the two methods. Asterisks in PCA score plots indicate group centroids.

MeOH models, 86, 85, and 92% were also recovered and classified as significant by the 95% EtOH models. Features of interest identified by the multivariate analysis were also investigated via univariate analysis. A high degree of correlation (R > 0.9) was observed for the feature Log2FC between subjects obtained via the two different extraction pipelines (Figure 2D). Focusing on annotated molecules of interest in microbiome research, such as primary and secondary bile acids, indole amino acids, vitamin B, heme catabolism, microbial *N*-acyl lipids, and long chain fatty acids, the 95% EtOH extraction

https://doi.org/10.1021/acs.analchem.4c05142
Anal. Chem. 2025, 97, 4945−4953

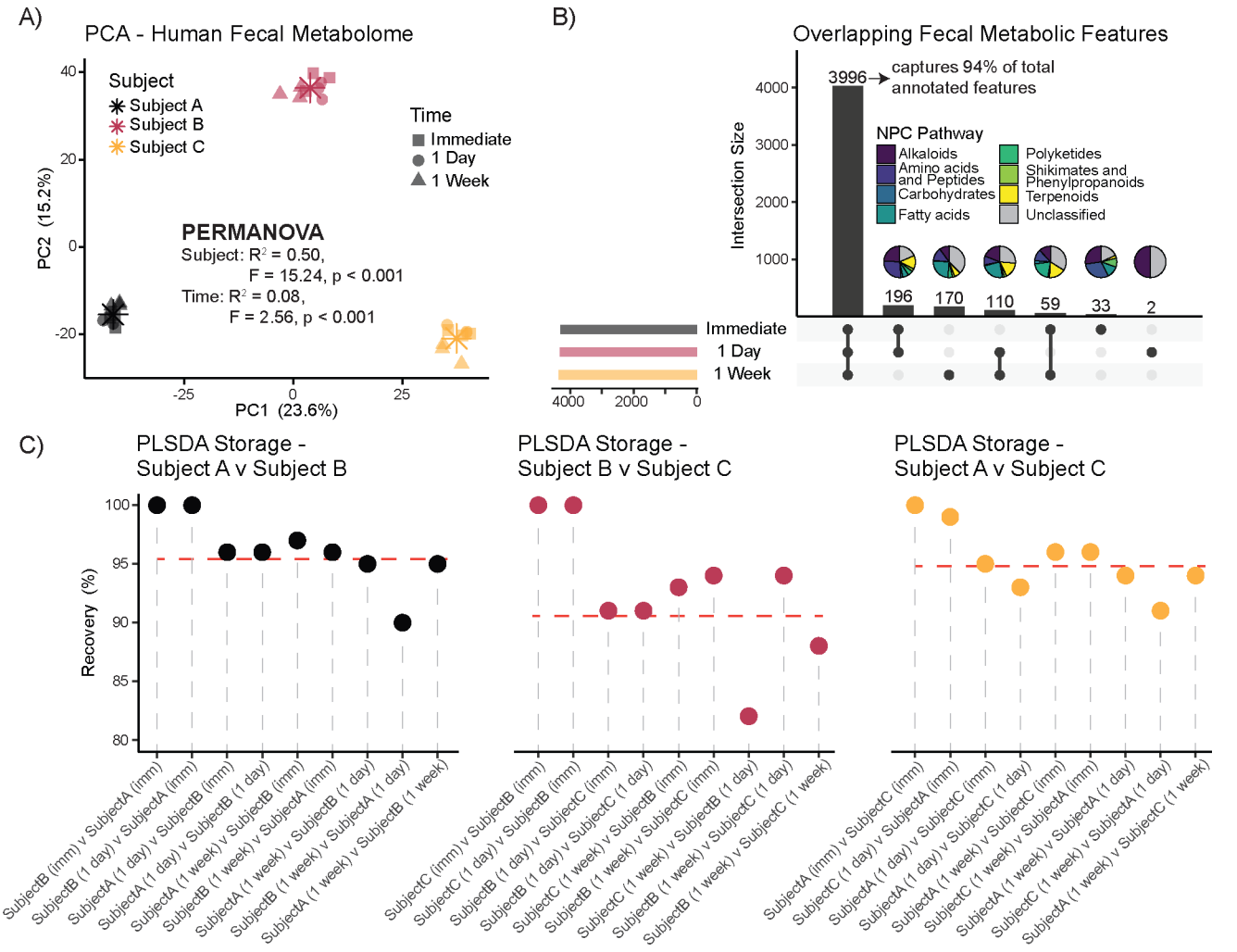


Figure 3. Fecal metabolome is relatively stable for up to 1 week at room temperature in 95% EtOH. (A) PCA on RCLR-transformed peak areas shows strong clustering based on subject id and little effect of storage time (PERMANOVA: subject $R^2 = 0.50$, $p < 0.001$ and extraction $R^2 = 0.08$, $p = 0.015$). (B) Upset plot displays that more than 88% of the obtained features are recovered via both extraction methods; these include 94% of the total annotated features. Most of the features exclusively characterizing the different storage time points are not annotated, and no class prediction can be obtained via CANOPUS. (C) Lollipop plots displaying the percentage of recovery of the top 100 features discriminating subjects via pairwise PLS-DA models. Immediate extractions (imm vs imm) are considered the "ground truth" (100%). Dashed lines represent the average percentage of recovery of the "ground truth" obtained via pairwise models constructed using different storage time points. Asterisks in PCA score plot indicate group centroids.

recovers and returns results comparable to those obtained via 50% MeOH extraction. Recovery examples of tri- and dihydroxylated bile acids, lysine, and histidine microbially conjugated bile acids, indoles such as tryptophan and propionic acid, vitamin B5, stercobilin, arginine-C5:0, and ω-3 arachidonic acid are available in Supplementary Figure 3.

Fecal Metabolome Remains Stable for Up to 1 Week at Room Temperature in 95% EtOH. PCA of triplicates exclusively extracted via the 95% EtOH pipeline showed clear clustering of samples based on subject ID (PERMANOVA, $R^2 = 0.50$, $F = 15.24$, $p < 0.001$) and a smaller effect of sample storage (PERMANOVA, $R^2 = 0.08$, $F = 2.56$, $p < 0.001$), which included immediate processing or storage at room temperature for either a day or a week in 95% EtOH (Figure 3A). Out of the 4566 features obtained via EtOH extraction, 88% (3996) were captured at all time points, comprising 94% (231) of all annotated features (Figure 3B). The majority of the features distinctively characterizing immediate (imm), week, immediate and day, and immediate and week extractions (156 out of 570) were not classified by CANOPUS (Supplementary Table 2). The subsequent most affected predicted classes were small peptides (65) and oligopeptides (48). Additionally, to determine if the storage time at room temperature could affect the metabolic features responsible for subject classification, 27 subject pairwise PLS-DA models were generated comparing samples of each subject at each different time point. Comparisons between immediate extractions were considered as "ground truths," and the relative recovery of those significant features was investigated for the different time points (Figure 3C). On average, 95.6, 91.6, and 94.7% of the top 100 "ground truth" features were recovered for each pairwise subject comparison. This suggested that the fecal metabolome could remain relatively stable in 95% EtOH for up to 1 week at room temperature without losing predictive power for subject discrimination. Log2FC in relation to storage time (immediate vs 1 week) of the metabolic were also explored for each subject (Supplementary Figure 4A). Interestingly, the median fold changes of the discriminating features of interest identified via

https://doi.org/10.1021/acs.analchem.4c05142
Anal. Chem. 2025, 97, 4945−4953

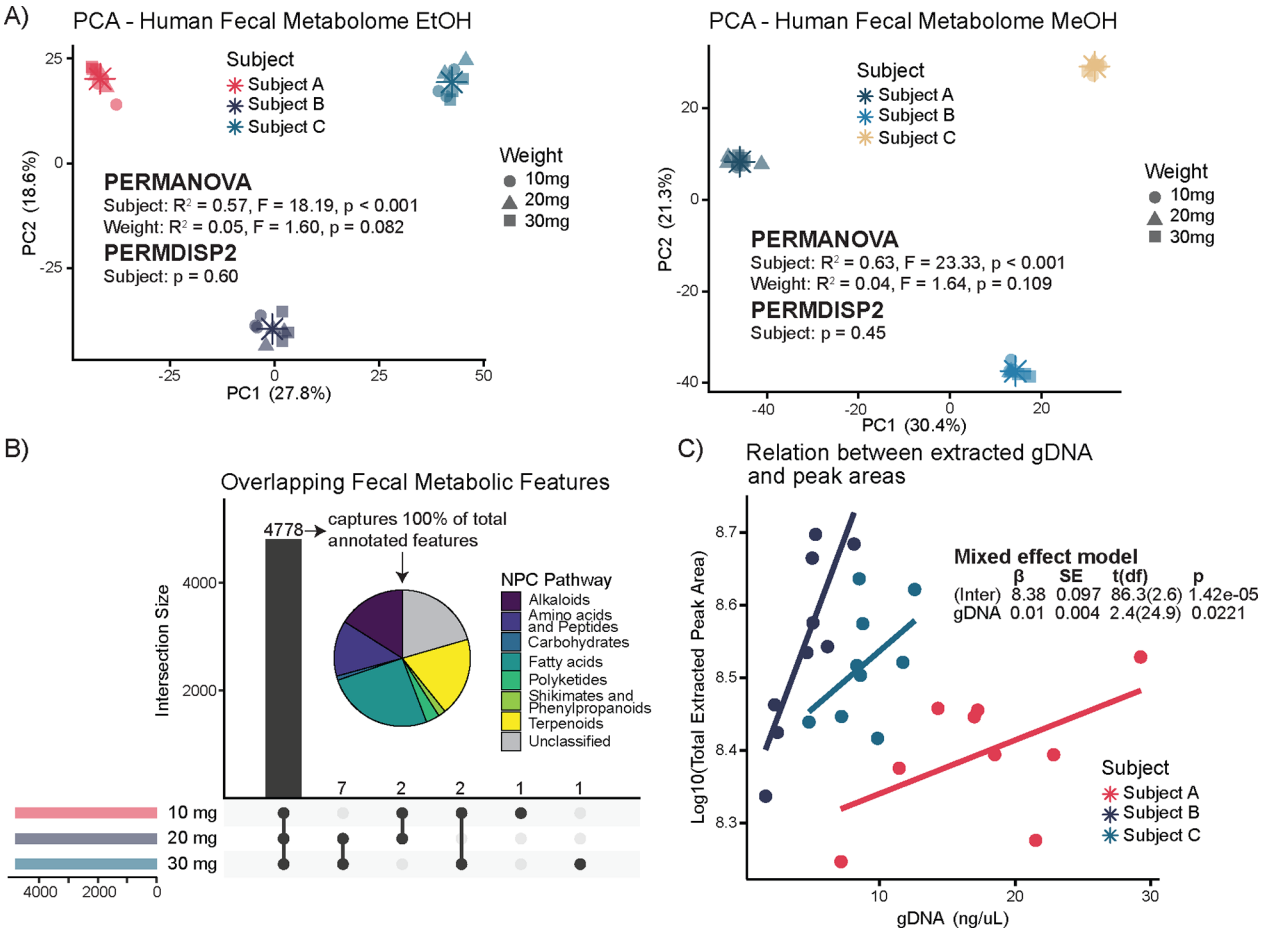


Figure 4. RCLR transformation in fecal untargeted metabolomics data. (A) PCA on RCLR-transformed peak areas removes sample weight variance in both 95% EtOH and 50% MeOH extractions (PERMANOVA: weight $R^2 < 0.05$, $p > 0.08$) and creates homogeneous variance for samples belonging to the same subject. (B) Upset plot displays that 99.9% of the features are captured by all of the different weight aliquots, also comprising all of the annotated features. Pie chart displays CANOPUS-predicted NPC pathways. (C) Scatter plot illustrates that extracted gDNA from the fecal samples is positively associated with the total extracted peak areas. Linear mixed effect model with subject id as random effect ($\beta = 0.010$, SE = 0.004, $p = 0.0221$). β = estimate; SE = standard error. Asterisks in PCA score plots indicate group centroids.

the pairwise PLS-DA models were 0.705 (Subject A vs Subject B), 0.430 (Subject B vs Subject C), and 0.530 (Subject A vs Subject C), suggesting a relative stability during storage time of the features of interest (Supplementary Figure 4B).

RCLR Transformation Can Address Discrepancies in Fecal Sample Collection. A recurrent issue in off-site fecal sample collection is the uneven sample amount collected by the individuals and the variability in the fecal water content. These can affect the per sample metabolite recovery and introduce bias into the downstream data analysis. Although lyophilization and weighting of the fecal samples can be implemented before data acquisition, these are time-consuming and unpractical when analyzing thousands of samples, and several technical errors can be introduced, such as repeated freeze−thaw cycles, intersample contamination, and others. To overcome these limitations, post data acquisition normalization methods can be deployed. Here, we investigate the use of RCLR transformation in fecal untargeted metabolomics data as a method to improve accuracy in data analysis. Originally introduced to tackle the microbiome data compositionality,[29] the RCLR transformation suits the semicompositionality nature of fecal untargeted metabolomics data. Additionally, RCLR allows for easy interpretability and seamless multiomics integration with associated RCLR-transformed microbiome data using tools such as Joint-RPCA[31] and DIABLO.[32] We investigated the extraction of different sample weights, 10, 20, and 30 mg, in both 95% EtOH and 50% MeOH extraction pipelines. PCA of RCLR data shows no effect of sample weight in both extraction methods (PERMANOVAs, $R^2 < 0.05$, $p > 0.08$), with most of the variance explained by subject id (R^2 = ~60%) and no difference in the dispersion of the samples (PERMDISP2, $p > 0.45$) between the different subjects (Figure 4A). On the contrary, PCA of raw data or relative abundance transformed data displayed less tightened clusters based on subjects, significant differences between group variances (PERMDISP2, $p < 0.001$), and small but still significant effect of the sample weight (Supplementary Figure 5). Upset plot revealed that the number of detected metabolites was not affected by sample weight, suggesting that just 10 mg is sufficient to cover the detectable fecal metabolome in reverse-phase LC-MS/MS data acquired in the positive ionization mode (Figure 4B). Finally, we explored correlations between weight, total obtained gDNA (genomic DNA, ng/μL), and cumulative extracted peak areas from samples processed via the Matrix Method (95% EtOH) and stored for 1 week at room temperature. Linear mixed effect models, accounting for repeated measures on the same subjects, found weight to be a significant predictor of total

https://doi.org/10.1021/acs.analchem.4c05142
Anal. Chem. 2025, 97, 4945−4953

extracted gDNA (β = 0.25, SE = 0.08, $t(23)$ = 2.92, p = 0.00768; Supplementary Figure 6A) and the total extracted peak areas (β = 0.01, SE = 0.001, $t(23)$ = 9.23, $p = 3.36 \times 10^{-9}$; Supplementary Figure 6B). Interestingly, gDNA recovery appeared also to be significantly correlated with the total extracted peak areas [β = 0.01, SE = 0.004, $t(24.98)$ = 2.441, p = 0.0221] but with variation in the slopes and intercepts between subjects (Figure 4C).

Limitations. This study focuses exclusively on human fecal samples, and it has not been validated for other types of biosamples. Data were acquired via reverse-phase liquid chromatography and positive ionization mode. For this reason, nondetected features might display a different behavior. Provided annotations are obtained via parent ion m/z and MS/MS spectral matching, resulting in a level 2 annotation according to the metabolomics standards initiative.[33] CANOPUS was used to generate molecular class predictions of unknown MS/MS spectra; as such, these predictions should be considered putative.

CONCLUSIONS

The presented study highlights that the recently introduced Matrix Method, which implements a 95% EtOH extraction, performs as well as a commonly used 50% MeOH extraction method for metabolomics analysis of human fecal samples. Moreover, the use of EtOH is safer compared to MeOH, and it can be handled by nonscientific personnel in the case of bulk off-site sample collections. Additionally, we showcase how the fecal metabolome remains relatively stable when stored in 95% EtOH for up to 1 week at room temperature, maintaining discriminating power between the investigated samples. This is important as samples cannot always immediately be stored at −80 °C, the most ideal condition, and refrigerated shipping can represent a high economic burden. Finally, we highlight how data analysis via RCLR transformation helps to remove variance possibly introduced by uneven sampling and discuss how this transformation suits better integrative microbiome studies. In conclusion, the 95% EtOH extraction of the Matrix Method represents a valid and more economically alternative to other widely used 50% MeOH extractions.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.4c05142.

PCA and PLS-DA score plots of pairwise subject comparisons, correlation between VIP scores obtained from pairwise PLS-DA models with either 50% MeOH or 95% EtOH extraction, boxplots of univariate analysis of annotated metabolites including bile acids, indoles, vitamins, and fatty acids, fold changes after storage for 1 week at room temperature, PCA score plots of raw data or data after relative abundance normalization for both 50% MeOH and 95% EtOH, correlation between sample weight and extracted gDNA or cumulative peak areas, and tables with NPC superclass predictions for extraction methods comparison (50% MeOH vs 95% EtOH) or storage comparison (immediate vs 1 week) (PDF)

AUTHOR INFORMATION

Corresponding Author

Simone Zuffa − *Skaggs School of Pharmacy and Pharmaceutical Sciences and Collaborative Mass Spectrometry Innovation Center, University of California San Diego, La Jolla, California 92093, United States;* orcid.org/0000-0001-7237-3402; Email: szuffa@health.ucsd.edu

Authors

Vincent Charron-Lamoureux − *Skaggs School of Pharmacy and Pharmaceutical Sciences and Collaborative Mass Spectrometry Innovation Center, University of California San Diego, La Jolla, California 92093, United States*

Caitriona Brennan − *Department of Pediatrics and Division of Biological Sciences, University of California San Diego, La Jolla, California 92093, United States*

Madison Ambre − *Department of Pediatrics, University of California San Diego, La Jolla, California 92093, United States*

Rob Knight − *Department of Pediatrics, Department of Computer Science and Engineering, Shu Chien-Gene Lay Department of Bioengineering, and Halıcıoğlu Data Science Institute, and Center for Microbiome Innovation, University of California San Diego, La Jolla, California 92093, United States*

Pieter C. Dorrestein − *Skaggs School of Pharmacy and Pharmaceutical Sciences and Collaborative Mass Spectrometry Innovation Center, University of California San Diego, La Jolla, California 92093, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.4c05142

Author Contributions

S.Z. designed the study, performed MeOH extraction and data analysis, and wrote the manuscript. V.C.L. performed MeOH extraction and ran the untargeted metabolomics experiments. C.B. provided fecal samples and designed and performed EtOH extraction. M.A. performed EtOH extraction. R.K. and P.C.D. secured fundings and provided supervision. All authors reviewed and approved the manuscript.

Notes

The authors declare the following competing financial interest(s): P.C.D. is an advisor and holds equity in Cybele, Sirenas, and BileOmix, and he is a scientific co-founder, advisor, and holds equity to Ometa, Enveda, and Arome with prior approval by UC San Diego. P.C.D. consulted for DSM Animal Health in 2023. R.K. is a scientific advisory board member, and consultant for BiomeSense, Inc., has equity and receives income. R.K. is a scientific advisory board member and has equity in GenCirq. R.K. is a consultant and scientific advisory board member for DayTwo, and receives income. R.K. has equity in and acts as a consultant for Cybele. R.K. is a co-founder of Biota, Inc., and has equity. R.K. is a co-founder and a scientific advisory board member of Micronoma, and has equity. The terms of these arrangements have been reviewed and approved by the University of California San Diego in accordance with its conflict of interest policies. All other authors declare no conflicts of interest.

https://doi.org/10.1021/acs.analchem.4c05142
Anal. Chem. 2025, 97, 4945−4953

REFERENCES

(1) Martino, C.; Dilmore, A. H.; Burcham, Z. M.; Metcalf, J. L.; Jeste, D.; Knight, R. *Nat. Rev. Microbiol.* **2022**, *20* (12), 707−720.

(2) Vuong, H. E.; Pronovost, G. N.; Williams, D. W.; Coley, E. J. L.; Siegler, E. L.; Qiu, A.; Kazantsev, M.; Wilson, C. J.; Rendon, T.; Hsiao, E. Y. *Nature* **2020**, *586* (7828), 281−286.

(3) Pronovost, G. N.; Yu, K. B.; Coley-O'Rourke, E. J. L.; Telang, S. S.; Chen, A. S.; Vuong, H. E.; Williams, D. W.; Chandra, A.; Rendon, T. K.; Paramo, J.; Kim, R. H.; Hsiao, E. Y. *Sci. Adv.* **2023**, *9* (40), No. eadk1887.

(4) Morton, J. T.; Jin, D.-M.; Mills, R. H.; Shao, Y.; Rahman, G.; McDonald, D.; Zhu, Q.; Balaban, M.; Jiang, Y.; Cantrell, K.; Gonzalez, A.; Carmel, J.; Frankiensztajn, L. M.; Martin-Brevet, S.; Berding, K.; Needham, B. D.; Zurita, M. F.; David, M.; Averina, O. V.; Kovtun, A. S.; Noto, A.; Mussap, M.; Wang, M.; Frank, D. N.; Li, E.; Zhou, W.; Fanos, V.; Danilenko, V. N.; Wall, D. P.; Cárdenas, P.; Baldeón, M. E.; Jacquemont, S.; Koren, O.; Elliott, E.; Xavier, R. J.; Mazmanian, S. K.; Knight, R.; Gilbert, J. A.; Donovan, S. M.; Lawley, T. D.; Carpenter, B.; Bonneau, R.; Taroncher-Oldenburg, G. *Nat. Neurosci.* **2023**, *26* (7), 1208−1217.

(5) Su, Q.; Wong, O. W. H.; Lu, W.; Wan, Y.; Zhang, L.; Xu, W.; Li, M. K. T.; Liu, C.; Cheung, C. P.; Ching, J. Y. L.; Cheong, P. K.; Leung, T. F.; Chan, S.; Leung, P.; Chan, F. K. L.; Ng, S. C. *Nat. Microbiol.* **2024**, *9*, 2344.

(6) Steimle, A.; Neumann, M.; Grant, E. T.; Willieme, S.; De Sciscio, A.; Parrish, A.; Ollert, M.; Miyauchi, E.; Soga, T.; Fukuda, S.; Ohno, H.; Desai, M. S. *Nat. Microbiol.* **2024**, *9*, 2244.

(7) Roje, B.; Zhang, B.; Mastrorilli, E.; Kovačić, A.; Sušak, L.; Ljubenkov, I.; Ćosić, E.; Vilović, K.; Meštrović, A.; Vukovac, E. L.; Bučević-Popović, V.; Puljiz, Ž.; Karaman, I.; Terzić, J.; Zimmermann, M. *Nature* **2024**, *632*, 1137.

(8) Bauermeister, A.; Mannochio-Russo, H.; Costa-Lotufo, L. V.; Jarmusch, A. K.; Dorrestein, P. C. *Nat. Rev. Microbiol.* **2022**, *20* (3), 143−160.

(9) Marotz, C.; Cavagnero, K. J.; Song, S. J.; McDonald, D.; Wandro, S.; Humphrey, G.; Bryant, M.; Ackermann, G.; Diaz, E.; Knight, R. *mSystems* **2021**, *6* (2), No. e0132920.

(10) Brennan, C.; Belda-Ferre, P.; Zuffa, S.; Charron-Lamoureux, V.; Mohanty, I.; Ackermann, G.; Allaband, C.; Ambre, M.; Boyer, T.; Bryant, M.; Cantrell, K.; Gonzalez, A.; McDonald, D.; Salido, R. A.; Song, S. J.; Wright, G.; Dorrestein, P. C.; Knight, R. *mSystems* **2024**, *9*, No. e0098524.

(11) Loftfield, E.; Vogtmann, E.; Sampson, J. N.; Moore, S. C.; Nelson, H.; Knight, R.; Chia, N.; Sinha, R. *Cancer Epidemiol. Biomarkers Prev.* **2016**, *25* (11), 1483−1490.

(12) Lim, M. Y.; Hong, S.; Kim, B.-M.; Ahn, Y.; Kim, H.-J.; Nam, Y.-D. *Sci. Rep.* **2020**, *10* (1), 1789.

(13) Ramamoorthy, S.; Levy, S.; Mohamed, M.; Abdelghani, A.; Evans, A. M.; Miller, L. A. D.; Mehta, L.; Moore, S.; Freinkman, E.; Hourigan, S. K. *BMC Microbiol.* **2021**, *21* (1), 59.

(14) Isokääntä, H.; Pinto da Silva, L.; Karu, N.; Kallonen, T.; Aatsinki, A.-K.; Hankemeier, T.; Schimmel, L.; Diaz, E.; Hyötyläinen, T.; Dorrestein, P. C.; Knight, R.; Orešič, M.; Kaddurah-Daouk, R.; Dickens, A. M.; Lamichhane, S. *Anal. Chem.* **2024**, *96* (22), 8893−8904.

(15) Han, S.; Van Treuren, W.; Fischer, C. R.; Merrill, B. D.; DeFelice, B. C.; Sanchez, J. M.; Higginbottom, S. K.; Guthrie, L.; Fall, L. A.; Dodd, D.; Fischbach, M. A.; Sonnenburg, J. L. *Nature* **2021**, *595* (7867), 415−420.

(16) Lai, Y.; Liu, C.-W.; Yang, Y.; Hsiao, Y.-C.; Ru, H.; Lu, K. *Nat. Commun.* **2021**, *12* (1), 6000.

(17) Gauglitz, J. M.; West, K. A.; Bittremieux, W.; Williams, C. L.; Weldon, K. C.; Panitchpakdi, M.; Di Ottavio, F.; Aceves, C. M.; Brown, E.; Sikora, N. C.; Jarmusch, A. K.; Martino, C.; Tripathi, A.; Meehan, M. J.; Dorrestein, K.; Shaffer, J. P.; Coras, R.; Vargas, F.; Goldasich, L. D.; Schwartz, T.; Bryant, M.; Humphrey, G.; Johnson, A. J.; Spengler, K.; Belda-Ferre, P.; Diaz, E.; McDonald, D.; Zhu, Q.; Elijah, E. O.; Wang, M.; Marotz, C.; Sprecher, K. E.; Vargas-Robles, D.; Withrow, D.; Ackermann, G.; Herrera, L.; Bradford, B. J.; Marques, L. M. M.; Amaral, J. G.; Silva, R. M.; Veras, F. P.; Cunha, T. M.; Oliveira, R. D. R.; Louzada-Junior, P.; Mills, R. H.; Piotrowski, P. K.; Servetas, S. L.; Da Silva, S. M.; Jones, C. M.; Lin, N. J.; Lippa, K. A.; Jackson, S. A.; Daouk, R. K.; Galasko, D.; Dulai, P. S.; Kalashnikova, T. I.; Wittenberg, C.; Terkeltaub, R.; Doty, M. M.; Kim, J. H.; Rhee, K. E.; Beauchamp-Walters, J.; Wright, K. P., Jr; Dominguez-Bello, M. G.; Manary, M.; Oliveira, M. F.; Boland, B. S.; Lopes, N. P.; Guma, M.; Swafford, A. D.; Dutton, R. J.; Knight, R.; Dorrestein, P. C. *Nat. Biotechnol.* **2022**, *40*, 1774.

(18) Dilmore, A. H.; Martino, C.; Neth, B. J.; West, K. A.; Zemlin, J.; Rahman, G.; Panitchpakdi, M.; Meehan, M. J.; Weldon, K. C.; Blach, C.; Schimmel, L.; Kaddurah-Daouk, R.; Dorrestein, P. C.; Knight, R.; Craft, S.; Alzheimer's Gut Microbiome Project Consortium. *Alzheimers Dement.* **2023**, *19* (11), 4805−4816.

(19) Quinn, R. A.; Melnik, A. V.; Vrbanac, A.; Fu, T.; Patras, K. A.; Christy, M. P.; Bodai, Z.; Belda-Ferre, P.; Tripathi, A.; Chung, L. K.; Downes, M.; Welch, R. D.; Quinn, M.; Humphrey, G.; Panitchpakdi, M.; Weldon, K. C.; Aksenov, A.; da Silva, R.; Avila-Pacheco, J.; Clish, C.; Bae, S.; Mallick, H.; Franzosa, E. A.; Lloyd-Price, J.; Bussell, R.; Thron, T.; Nelson, A. T.; Wang, M.; Leszczynski, E.; Vargas, F.; Gauglitz, J. M.; Meehan, M. J.; Gentry, E.; Arthur, T. D.; Komor, A. C.; Poulsen, O.; Boland, B. S.; Chang, J. T.; Sandborn, W. J.; Lim, M.; Garg, N.; Lumeng, J. C.; Xavier, R. J.; Kazmierczak, B. I.; Jain, R.; Egan, M.; Rhee, K. E.; Ferguson, D.; Raffatellu, M.; Vlamakis, H.; Haddad, G. G.; Siegel, D.; Huttenhower, C.; Mazmanian, S. K.; Evans, R. M.; Nizet, V.; Knight, R.; Dorrestein, P. C. *Nature* **2020**, *579* (7797), 123−129.

(20) Gentry, E. C.; Collins, S. L.; Panitchpakdi, M.; Belda-Ferre, P.; Stewart, A. K.; Carrillo Terrazas, M.; Lu, H.-H.; Zuffa, S.; Yan, T.; Avila-Pacheco, J.; Plichta, D. R.; Aron, A. T.; Wang, M.; Jarmusch, A. K.; Hao, F.; Syrkin-Nikolau, M.; Vlamakis, H.; Ananthakrishnan, A. N.; Boland, B.; Hemperly, A.; Vande Casteele, N.; Gonzalez, F. J.; Clish, C. B.; Xavier, R. J.; Chu, H.; Baker, E. S.; Patterson, A. D.; Knight, R.; Siegel, D.; Dorrestein, P. C. *Nature* **2024**, *626*, 419.

(21) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. *Nat. Biotechnol.* **2012**, *30* (10), 918−920.

(22) Schmid, R.; Heuckeroth, S.; Korf, A.; Smirnov, A.; Myers, O.; Dyrlund, T. S.; Bushuiev, R.; Murray, K. J.; Hoffmann, N.; Lu, M.; Sarvepalli, A.; Zhang, Z.; Fleischauer, M.; Dührkop, K.; Wesner, M.; Hoogstra, S. J.; Rudt, E.; Mokshyna, O.; Brungs, C.; Ponomarov, K.; Mutabdžija, L.; Damiani, T.; Pudney, C. J.; Earll, M.; Helmer, P. O.; Fallon, T. R.; Schulze, T.; Rivas-Ubach, A.; Bilbao, A.; Richter, H.; Nothias, L.-F.; Wang, M.; Orešič, M.; Weng, J.-K.; Böcker, S.; Jeibmann, A.; Hayen, H.; Karst, U.; Dorrestein, P. C.; Petras, D.; Du, X.; Pluskal, T. *Nat. Biotechnol.* **2023**, *41* (4), 447−449.

(23) Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; Aicheler, F.; Aksenov, A. A.; Alka, O.; Allard, P.-M.; Barsch, A.; Cachet, X.; Caraballo-Rodriguez, A. M.; Da Silva, R. R.; Dang, T.; Garg, N.; Gauglitz, J. M.; Gurevich, A.; Isaac, G.; Jarmusch, A. K.; Kameník, Z.; Kang, K. B.; Kessler, N.; Koester, I.; Korf, A.; Le Gouellec, A.; Ludwig, M.; Martin, H. C.; McCall, L.-I.; McSayles, J.; Meyer, S. W.; Mohimani, H.; Morsy, M.; Moyne, O.; Neumann, S.; Neuweger, H.; Nguyen, N. H.; Nothias-Esposito, M.; Paolini, J.; Phelan, V. V.; Pluskal, T.; Quinn, R. A.; Rogers, S.; Shrestha, B.; Tripathi, A.; van der Hooft, J. J. J.; Vargas, F.; Weldon, K. C.; Witting, M.; Yang, H.; Zhang, Z.; Zubeil, F.; Kohlbacher, O.; Böcker, S.; Alexandrov, T.; Bandeira, N.; Wang, M.; Dorrestein, P. C. *Nat. Methods* **2020**, *17* (9), 905−908.

https://doi.org/10.1021/acs.analchem.4c05142
Anal. Chem. 2025, 97, 4945−4953

(24) Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Böcker, S. *Nat. Methods* **2019**, *16* (4), 299−302.
(25) Kirwan, J. A.; Gika, H.; Beger, R. D.; Bearden, D.; Dunn, W. B.; Goodacre, R.; Theodoridis, G.; Witting, M.; Yu, L.-R.; Wilson, I. D.; metabolomics Quality Assurance and Quality Control Consortium (mQACC). *Metabolomics* **2022**, *18* (9), 70.
(26) Mildau, K.; van der Hooft, J. J. J.; Flasch, M.; Warth, B.; El Abiead, Y.; Koellensperger, G.; Zanghellini, J.; Büschl, C. *Bioinformatics* **2022**, *38* (22), 5139−5140.
(27) Kuhn, M. *J. Stat. Softw.* **2008**, *28*, 1−26.
(28) Rohart, F.; Gautier, B.; Singh, A.; Lê Cao, K.-A. *PLoS Comput. Biol.* **2017**, *13* (11), No. e1005752.
(29) Martino, C.; Morton, J. T.; Marotz, C. A.; Thompson, L. R.; Tripathi, A.; Knight, R.; Zengler, K. *mSystems* **2019**, *4* (1), No. e0001619.
(30) Conway, J. R.; Lex, A.; Gehlenborg, N. *Bioinformatics* **2017**, *33* (18), 2938−2940.
(31) Burcham, Z. M.; Belk, A. D.; McGivern, B. B.; Bouslimani, A.; Ghadermazi, P.; Martino, C.; Shenhav, L.; Zhang, A. R.; Shi, P.; Emmons, A.; Deel, H. L.; Xu, Z. Z.; Nieciecki, V.; Zhu, Q.; Shaffer, M.; Panitchpakdi, M.; Weldon, K. C.; Cantrell, K.; Ben-Hur, A.; Reed, S. C.; Humphry, G. C.; Ackermann, G.; McDonald, D.; Chan, S. H. J.; Connor, M.; Boyd, D.; Smith, J.; Watson, J. M. S.; Vidoli, G.; Steadman, D.; Lynne, A. M.; Bucheli, S.; Dorrestein, P. C.; Wrighton, K. C.; Carter, D. O.; Knight, R.; Metcalf, J. L. *Nat. Microbiol* **2024**, *9* (3), 595−613.
(32) Singh, A.; Shannon, C. P.; Gautier, B.; Rohart, F.; Vacher, M.; Tebbutt, S. J.; Lê Cao, K.-A. *Bioinformatics* **2019**, *35* (17), 3055−3062.
(33) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.-M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reily, M. D.; Thaden, J. J.; Viant, M. R. *Metabolomics* **2007**, *3* (3), 211−221.

https://doi.org/10.1021/acs.analchem.4c05142
Anal. Chem. 2025, 97, 4945−4953