analytical chemistry

Editors' Highlight

pubs.acs.org/ac

Development of High-Performance Chemical Isotope Labeling LC–MS for Profiling the Human Fecal Metabolome

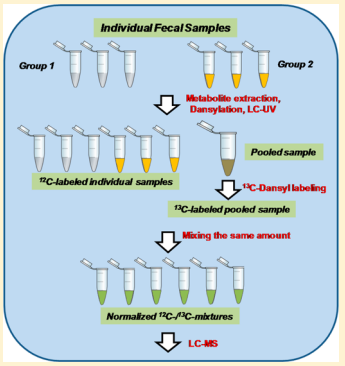
Wei Xu,[†] Deying Chen,[†] Nan Wang,[†,‡] Ting Zhang,[†] Ruokun Zhou,[‡] Tao Huan,[‡] Yingfeng Lu,[†] Xiaoling Su,[†] Qing Xie,[†] Liang Li,*[,†,‡] and Lanjuan Li*[,†]

[†]State Key Laboratory and Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou 310003, China

[‡]Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada

Supporting Information

ABSTRACT: Human fecal samples contain endogenous human metabolites, gut microbiota metabolites, and other compounds. Profiling the fecal metabolome can produce metabolic information that may be used not only for disease biomarker discovery, but also for providing an insight about the relationship of the gut microbiome and human health. In this work, we report a chemical isotope labeling liquid chromatography–mass spectrometry (LC–MS) method for comprehensive and quantitative analysis of the amine- and phenol-containing metabolites in fecal samples. Differential $^{13}C_2/^{12}C_2$-dansyl labeling of the amines and phenols was used to improve LC separation efficiency and MS detection sensitivity. Water, methanol, and acetonitrile were examined as an extraction solvent, and a sequential water–acetonitrile extraction method was found to be optimal. A step-gradient LC–UV setup and a fast LC–MS method were evaluated for measuring the total concentration of dansyl labeled metabolites that could be used for normalizing the sample amounts of individual samples for quantitative metabolomics. Knowing the total concentration was also useful for optimizing the sample injection amount into LC–MS to maximize the number of metabolites detectable while avoiding sample overloading. For the first time, dansylation isotope labeling LC–MS was performed in a simple time-of-flight mass spectrometer, instead of high-end equipment, demonstrating the feasibility of using a low-cost instrument for chemical isotope labeling metabolomics. The developed method was applied for profiling the amine/phenol submetabolome of fecal samples collected from three families. An average of 1785 peak pairs or putative metabolites were found from a 30 min LC–MS run. From 243 LC–MS runs of all the fecal samples, a total of 6200 peak pairs were detected. Among them, 67 could be positively identified based on the mass and retention time match to a dansyl standard library, while 581 and 3197 peak pairs could be putatively identified based on mass match using MyCompoundID against a Human Metabolome Database and an Evidence-based Metabolome Library, respectively. This represents the most comprehensive profile of the amine/phenol submetabolome ever detected in human fecal samples. The quantitative metabolome profiles of individual samples were shown to be useful to separate different groups of samples, illustrating the possibility of using this method for fecal metabolomics studies.



There is a growing interest in linking the activities of symbiotic gut microbes with human health.[1,2] This is mainly driven by recent advances in genome sequencing technologies which allow the analysis and characterization of genomic information on microbes very quickly.[3,4] For example, the Human Microbiome Project (HMP) funded by the U.S. National Institutes of Health (NIH) has generated a wealthy array of information on structure, function, and diversity of the human microbiome from healthy individuals with specimens collected in several body areas.[1,5,6] For this project, a stool or fecal specimen was also collected to represent the microbiota of the lower gastrointestinal tract.[6] Several studies of the human microbiome of feces and their potential association with human health have been recently reported.[7−9] While genome-association studies will undoubtedly continue, metabolomics represents another important area where the metabolomic information may provide direct structural and functional evidence on the intrinsic relations of microbes and human, which should lead to better understanding of the roles of the human microbiome on human health. In this regard, human feces are an excellent source to interrelate the microbiome and human metabolomes,[7,10,11] as the human fecal metabolome is comprised of endogenous human metabolites, gut microbiota metabolites, and residues or metabolites of digested materials.

There are a growing number of studies reported on fecal metabolomics of human and animal models.[12−16] These studies have demonstrated that fecal metabolome profiling can reveal the significant metabolites differentiating comparative groups. Fecal metabolome profiles have been showed to be useful to differentiate diseased group vs control group, such as in the studies of ulcerative colitis and irritable bowel syndrome,[17] liver cirrhosis and hepatocellular carcinoma,[18] chronic kidney

Received: September 26, 2014
Accepted: December 8, 2014
Published: December 8, 2014

   dx.doi.org/10.1021/ac503619q | Anal. Chem. 2015, 87, 829–836

disease,[19] and intestinal transplant rejection.[20] Because of the important role of gut microbes on food digestion, fecal metabolomics has been used to assess the impact of various types of food on the human microbiome and human health.[21,22] Other reported studies include using fecal metabolomics to study the aging process.[23] One recent study showed that fecal metabolome profiling of conventional mice and humanized mice (i.e., mice colonized with a human microbiota) could be used to investigate the role of human-relevant microbes on host metabolomes.[11] The use of an human-relevant animal model opens the possibility of applying fecal metabolomics to study the gut microbiota functions related to human health under a well-controlled experimental setting.

Almost all the fecal metabolome profiling experiments reported so far were done using NMR, gas chromatography−mass spectrometry (GC−MS) and, to a less extent, LC−MS. These techniques provided varying degrees of metabolome coverage. Unlike other human biofluids such as urine and blood, fecal samples are more difficult to deal with, although they can be readily obtained noninvasively. There are reports of method development for NMR and GC−MS based metabolome analysis, including the studies of optimal sample extraction, NMR, or GC−MS experimental protocols.[24−26] Reversed-phase (RP) and hydrophilic interaction (HILIC) columns, in conjunction with the use of both positive and negative ion modes of detection, have been used for LC−MS analysis of mouse fecal samples.[11] However, because of the great diversity of metabolites and limited analytical capability of each method, there is still a great need to increase the metabolome coverage in fecal metabolome profiling.

Recently, we have shown that chemical isotope labeling (CIL) LC−MS using a rationally designed labeling reagent can increase the metabolite detectability significantly.[27,28] For example, $^{13}C_2/^{12}C_2$-dansyl reagents can be used to label amine- and phenol-containing metabolites and the resultant labeled metabolites can be efficiently separated using RPLC and detected by electrospray ionization (ESI) MS with 10- to 1000-fold increase in detection sensitivity.[27] In a dansyl labeled human urine sample, over 20 000 features can be detected in a 25 min LC−MS run, resulting in the identification of more than 1 600 unique peak pairs or putative metabolites.[29] In this paper, we report a method for the analysis of the human fecal submetabolome based on differential isotope dansyl labeling LC−MS and demonstrate a significant improvement in amine/phenol submetabolome analysis. Although only amines and phenols, which are important groups of metabolites found in many metabolic pathways, are targeted in this work, the method developed should be applicable to other groups of metabolites using different labeling chemistries. We also illustrate, for the first time, that a simple time-of-flight (TOF) mass spectrometer with a resolving power of about 15 000 is well-suited for CIL LC−MS; previous work was done using Fourier transform ion cyclotron resonance (FTICR)-MS or high-end quadrupole-time-of-flight (QTOF)-MS.[30,31] We apply this method to profile the fecal submetabolomes of three families (parents plus an infant per family with daily sample collection for 3 days) and illustrate the applicability of this method to generate quantitative metabolome information to differentiate different groups.

EXPERIMENTAL SECTION

Overall Workflow. Figure 1 shows the general workflow for fecal metabolome profiling using CIL LC−MS. Water is first added to a fecal sample, followed by vortexing to produce a homogenized suspension. An aliquot of the suspension is taken and dried using a SpeedVac. The dried sample is subjected to solvent extraction. After centrifugation, the supernatant is taken for ^{12}C-dansylation labeling. The labeled sample is injected into LC−UV in order to determine the total concentration of labeled metabolites in an individual sample. On the basis of the concentration, an appropriate sample volume is taken to generate a pooled sample. The same amount of individual samples is used for pooling to ensure equal representation. This pooled sample is then labeled by ^{13}C-dansylation, which serves as a control or a global internal standard. To perform relative metabolite quantification of individual samples, an equal amount of the ^{12}C-labeled individual sample is mixed with the ^{13}C-labeled pooled sample. Since the same pooled sample is used as a control for all the individual samples, the mass spectral peak ratios of a given metabolite can be used to measure the concentration differences in the individual samples. Because the pooled sample and an individual sample are processed in the same manner using differential isotope labeling, accurate and precise relative quantification results can be generated.[27]

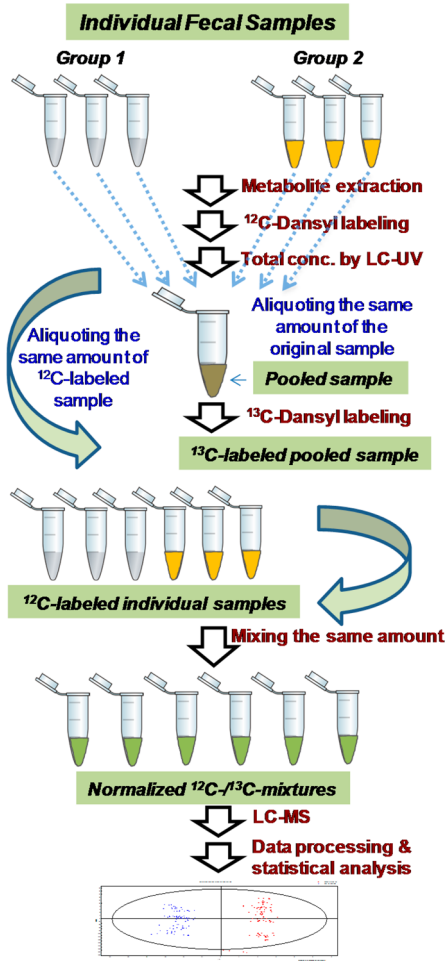


Figure 1. Workflow of the differential chemical isotope labeling LC−MS method for fecal metabolomics.

Fecal Sample Collection. All samples were collected in compliance with prevailing human research ethics guidelines. The study protocol was approved by the Medical Ethics Committee of the first Affiliated Hospital of Zhejiang

dx.doi.org/10.1021/ac503619q | Anal. Chem. 2015, 87, 829−836

University, Hangzhou, China (No. 2014-345) and the Ethics Approval Board of the University of Alberta, Edmonton, Canada (No. MS8_Pro00012946). Three Chinese families (two parents and one infant per family) with no known diseases participated in this study. Their ages were 29 years old (male), 27 (female), and 1 month (infant) for family no. 1; 29 (male), 27 (female), and 2 months (infant) for family no. 2; and 29 (male), 28 (female), and 1 month (infant) for family no. 3. For the adults, stool was collected daily in the morning after 12-h fasting in three separate days. For the infants, the time of collecting the daily stool was not controlled. Fresh stools were set at room temperature for less than 2 h after collection and were stored at −80 °C.

Fecal Sample Processing. Fecal samples collected daily from three families (father, mother, and infant per family) in three separate days were stored at −80 °C and then thawed to room temperature for processing. Water was first added to a sample, followed by vortexing to produce a homogenized solution or suspension from which 15 1 mL aliquots were taken. The aliquots were dried using a SpeedVac. The dried samples were subjected to solvent extraction. Water, methanol (MeOH), and acetonitrile (ACN) were selected for extraction. In each case of solvent extraction, three dried aliquots of a fecal sample were taken for experimental triplicates. Each dried aliquot was extracted 3 times with 100 μL of solvent for each extraction by ultrasonication for 10 min (KH-250B, Kunshan, Jiangshu, China), vortexing for 2 min, and then centrifugation at 18 000*g* for 15 min at 4 °C. The supernatants from the three extractions were combined. The combined extracts were dried with a SpeedVac and then stored at −20 °C for further use. A sequential solvent extraction method in the order of water and ACN (H_2O → ACN) was also evaluated. In this case, 150 μL of water was added into a dried aliquot for the first extraction, followed by using 150 μL of ACN for the second extraction. The supernatants from the two extractions were combined and dried with a SpeedVac, then stored at −20 °C.

Dansylation Labeling. Dansylation labeling was performed according to a previously reported protocol[27,31] and detailed information is provided in Supplemental Note N1 in the Supporting Information.

LC−UV. The ^{12}C-labeled individual samples were separately injected onto LC−UV for quantifying the total labeled metabolites in each sample based on absorption at 338 nm.[32] An Agilent 1290 UHPLC system with a photodiode array detector (Agilent, Palo Alto, CA) was used for LC−UV. The column used was Waters ACQUITY UPLC BEH C18 column (2.1 mm × 10 cm, 1.7 μm particle size, 130 Å pore size). LC solvent A was 0.1% (v/v) formic acid in water, and solvent B was 0.1% (v/v) formic acid in ACN. The fast step-gradient elution profile was as follows: t = 0 min, 15% B; t = 1.00 min, 15% B; t = 1.01 min, 98% B; t = 2.00 min, 98% B; t = 2.50 min, 15% B; t = 6.00 min, 15% B. The flow rate was 500 μL/min, and the sample injection volume was 5 μL, unless otherwise stated.

LC−MS. An Agilent 1290 series binary UHPLC system with a Waters ACQUITY UPLC BEH C18 column (2.1 mm × 10 cm, 1.7 μm particle size, 130 Å pore size) connected to an Agilent electrospray ionization (ESI) time-of-flight mass spectrometer (6230, Agilent, Palo Alto, CA) was used for LC−MS. Supplemental Note N1 in the Supporting Information provides details of the LC and MS conditions used in this work.

Data Processing and Analysis. A software tool, IsoMS,[33] was used to process the raw data generated from multiple LC−MS runs by peak picking, peak pairing, peak-pair filtering, and peak-pair intensity ratio calculation. The same peak pairs detected from multiple samples were then aligned to produce a CSV file that contains the metabolite information and peak ratios relative to a control (i.e., a pooled sample). Some peak ratio values were missing in some runs. A zero-fill program[34] was used to find the missing peak pairs from the raw mass spectral data and fill in the missing values. The final metabolite-intensity data file was then exported to SIMCA-P+ 12.0 software (Umetrics, Umeå, Sweden) for multivariate statistical analysis. Principal component analysis (PCA) and orthogonal projections to latent structures discriminant analysis (OPLS-DA) were used to analyze the data. Metabolite identification was performed based on mass and retention time match to a dansyl standard library. Putative identification was done based on accurate mass match to the metabolites in the human metabolome database (HMDB) (8 021 known human endogenous metabolites) and the Evidence-based Metabolome Library (EML) (375 809 predicted human metabolites with one reaction) using MyCompoundID.[35] The mass accuracy tolerance window was set at 10 ppm for database search.

RESULTS AND DISCUSSION

The main objective of developing a workflow tailored to fecal metabolome analysis was to optimize the key steps involved (see Figure 1) in order to detect and quantify as many metabolites as possible. The rationale and performance of the method developed for each step are described below.

Solvent Extraction. Analyzing the fecal metabolome requires a different sample preparation protocol from those used to deal with homogeneous biological samples such as urine and blood. Effective extraction of the metabolites present in a fecal sample is very important.[25,26] Currently, there is no unified approach for solvent extraction; the optimized extraction method is dependent on the analytical technique used. Our goal was to develop a simple method that could be used to process many samples quickly. We first examined the effect of solvent type on the extraction process. Three solvents, i.e., water, methanol (MeOH), and acetonitrile (ACN), were selected, after considering their different solubility for dissolving amines and phenols.

In the study of solvent extraction methods, for each solvent, three aliquots of a fecal sample were taken for experimental triplicate. For each aliquot, after extraction, the extract was divided into two halves for ^{12}C and ^{13}C labeling separately. The labeled samples were mixed and then injected into the LC−MS for three replicate runs (i.e., technical triplicate). Thus, in total, 9 sets of LC−MS data were generated from each solvent extraction method. Supplemental Note N2 in the Supporting Information provides information on peak pair comparison of the data sets. As Supplemental Note N2 in the Supporting Information shows, the results of technical triplicate runs of a labeled aliquot were very reproducible. For example, 997, 985, and 988 peak pairs were detected from aliquot 1 of the water extract with 940 common pairs. In addition, many common peak pairs were detected from the experimental triplicate of an extract. For example, out of a total of 1106 pairs detected from the three aliquots of water extracts, 972 pairs (88%) were in common. These results indicate that within a solvent extract the number of peak pairs detected was reproducible.

dx.doi.org/10.1021/ac503619q | *Anal. Chem.* 2015, 87, 829−836

When the total number of peak pairs detected from an aliquot of water, MeOH, and ACN extracts was compared, we found that many more unique peak pairs were detected in water and ACN extracts, compared to MeOH extract. For example, in the aliquot 1 data set (see Supplemental Note N2 in the Supporting Information), 1037, 940, and 978 peak pairs were detected in the water, methanol, and ACN extracts, respectively, for a combined total of 1184 pairs. However, only 7 pairs were detected uniquely from the MeOH extract, representing less than 0.6% of the total number. Thus, the amine/phenol metabolites extracted from a fecal sample using MeOH could be mostly covered by the water and ACN extracts. On the basis of these results, we decided to evaluate a sequential solvent extraction method in the order of water and ACN (H_2O → ACN) to see if we could extract more metabolites than using either water or ACN alone.

In the H_2O → ACN extract, the number of peak pairs detected in triplicate injections was 1249, 1247, and 1247 from aliquot 1 (1196 common pairs), 1206, 1197, and 1202 from aliquot 2 (1168 common pairs), and 1251, 1256, and 1254 from aliquot 3 (1213 common pairs). An average of 1234 ± 25 (n = 9) pairs were detected per run from the H_2O → ACN extract, compared to 989 ± 6 (n = 9) from the water extract and 942 ± 14 (n = 6) from the ACN extract. Thus, more peak pairs per run were detected from the H_2O → ACN extract.

Supplemental Note N2 in the Supporting Information describes the comparison of the peak pairs detected among the water, ACN, and H_2O → ACN extracts. A large fraction of the peak pairs detected in the water or ACN extract could be detected in the H_2O → ACN extract. For example, in the aliquot 1 data set (see Supplemental Note N2 in the Supporting Information), 821 out of a total of 1398 peak pairs detected in all three solvent extracts (i.e., 59%) were in common, and 1298 or 93% pairs were detected in the H_2O → ACN extract. Only 28 unique peak pairs (2%) were found in the water extract and 89 unique pairs (6%) were detected in the ACN extract. These results indicated that H_2O → ACN sequential extraction performed better than using water or ACN alone. Thus, we used this sequential extraction method for the subsequent experiments.

Besides the number of peak pairs detected, quantitative results of the peak pairs are also important for metabolome profiling. We examined the peak ratio data generated from the H_2O → ACN extract. For relative quantification, the peak ratio within a peak pair of a differentially labeled metabolite is used in the CIL LC−MS workflow. In the 9 data sets generated from the H_2O → ACN extract, the average peak ratio of a data set was found to be 1.05 ± 0.13, very close to the expected ratio of 1.00. The commonly detected peak pairs in triplicate runs of an extract had an average coefficient variation (CV) of 3.9% with 99% peak pairs having a CV of <20% for aliquot 1, 3.8% with 99% with CV of <20% for aliquot 2, and 3.9% with 98% with CV of <20% for aliquot 3. These results indicate that good quantitative accuracy and reproducibility could be obtained. From the above studies, we concluded that H_2O → ACN extraction is a simple method to extract amine/phenol metabolites for dansylation LC−MS analysis.

Sample Amount Normalization. For relative metabolite quantification in a metabolomics work, it is very important to normalize the sample amount used for all the comparative samples. This is a major challenge for fecal samples, as they contain both liquid and solid materials with varying proportions and densities. The liquid content and solid density affect the metabolite concentration in a given sample weight or volume. Instead of taking an equal volume or weight of individual samples for comparative analysis, we propose to use LC−UV to measure the total amount of labeled metabolites in a sample and then use it to normalize all the individual samples.

Using LC−UV for sample amount normalization was possible, because dansylation labeling offers the possibility of performing UV quantification of the labeled metabolites. The dansyl group absorbs at around 338 nm where there are no significant interferences from other absorbers in a fecal extract. Supplemental Figure S1A in the Supporting Information shows the LC−UV chromatograms of a series of diluted solutions prepared from a stock solution of a labeled fecal sample. A fast step-gradient was used to separate the hydrolyzed dansyl product from the leftover reagent (eluted at high water content) and the dansyl labeled metabolites (eluted at high organic solvent content). Supplemental Figure S1B in the Supporting Information shows the LC−UV chromatogram of the unlabeled fecal sample. Triplicate experiments showed that the absorption of the unlabeled sample at 338 nm is less than 4% of the labeled metabolites, suggesting that the labeled metabolites could be used for quantification without much interference from the unlabeled metabolites.

Figure 2A shows the calibration curves generated by plotting the peak area of the labeled metabolites as a function of the fecal sample or amino acid concentration. It should be noted that the concentration of the stock solution used to prepare the

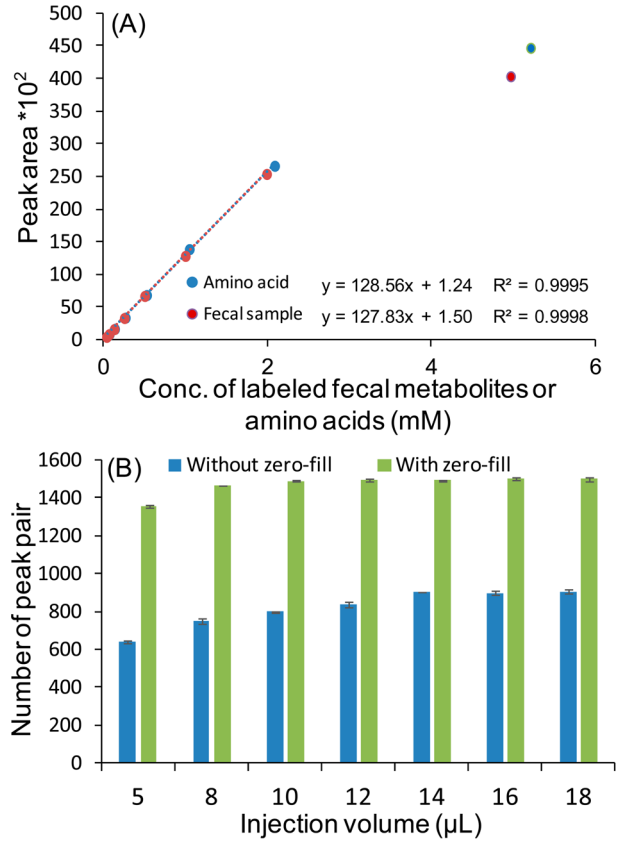


Figure 2. (A) LC−UV calibration curves of dansyl labeled fecal sample and amino acids used for measuring the total concentration of labeled metabolites (triplicate experiments). (B) Number of peak pairs detected in H_2O → ACN extract as a function of sample injection volume. The concentration of the labeled fecal sample was 4.96 mM.

dx.doi.org/10.1021/ac503619q | Anal. Chem. 2015, 87, 829−836

diluted solutions was determined by using a calibration curve established from a mixture of 17-dansyl amino acid standards (17-Dns-aas) (see Figure 2A). This is because a fecal sample with known concentrations of all the metabolites present is not available. This approach assumed that the absorptivity of the 17-amino-acid mixture is the same as that of the labeled metabolites in a fecal sample, which may not be true and thus the fecal concentration values shown in Figure 2A may not accurately represent the true values of the metabolite concentrations. However, knowing the absolute total metabolite concentration is not essential for relative metabolome comparison. We only need to control the sample volume so that we can sample the nominally same amount from each sample for analysis. Thus, we used the fecal sample calibration curve shown in Figure 2A to determine the total concentration of labeled metabolites in an individual fecal sample; for a sample of >2 mM, dilution of the sample to the linear range is needed.

We recognized that not all users would have access to a LC−UV instrument for sample amount normalization. We thus investigated the use of LC−MS, with a shorter gradient run than the normal LC−MS profiling run, as an alternative method to determine the total amount of labeled metabolites in a fecal sample. Supplemental Figure S2 in the Supporting Information shows the calibration curves of 17-Dns-aas and the labeled fecal sample obtained by plotting the total peak area of a fast LC−MS ion chromatogram (i.e., integrating all the peaks from 1.7 to 2.0 min) as a function of the analyte concentration. Compared to UV detection, MS signal response was not linear when the 17-Dns-aas concentration was above 0.065 mM. In the lower concentration range (<0.065 mM), a linear response was obtained (see Supplemental Figure S2C in the Supporting Information). Thus, it should be possible to normalize the sample amount using LC−MS, but high dilution of a labeled sample is needed in order to adjust the fecal sample concentration to fall within the linear range (see Supplemental Figure S2D in the Supporting Information). The major disadvantage of using LC−MS for sample normalization is that valuable MS instrument time is required for the LC−MS-based method. In our work, since we have an LC−UV system available, we chose the use of this less expensive method for sample amount normalization.

In profiling the metabolomes of fecal samples of three families, after homogenizing a sample dissolved in water, three equal aliquots were taken for triplicate experiments. Each aliquot was dried and then extracted using $H_2O \rightarrow$ ACN, followed by ^{12}C-dansylation and LC−UV quantification. The quantification results of 81 extract samples (triplicates of 27 fecal samples) are shown in Supplemental Table T1 in the Supporting Information. The total concentration of labeled metabolites in the extracts ranges from 0.60 to 6.37 mM; the original fecal extract concentration was 8.4-fold higher, taking into account the dilution factor during labeling and injection. These quantification results suggest that sample-to-sample variation in terms of total metabolite concentration can be quite large. Even for the samples collected from one individual at three different days, the total concentration of labeled metabolites can vary by more than 3-fold. This underscores the importance of performing sample amount normalization for relative quantification of the fecal metabolomes.

LC−MS Optimization. Both the LC and MS settings can affect the peak detectability. In this work, the LC separation conditions used was based on the optimized conditions previously reported.[27] Within the MS settings, the mass spectral acquisition rate can affect peak detectability. The TOF-MS used in this work has the lowest scan rate at 0.87 Hz. By injecting the same amount of sample from a 2.75 mM fecal solution, but varying the data acquisition rate, 373 ± 15 (n = 9), 486 ± 30, 522 ± 22, and 764 ± 23 peak pairs were detected at a rate of 4, 3, 2, and 1 Hz, respectively. These results indicate that 1 Hz is the optimal condition for data acquisition for this LC−TOF-MS instrument.

Besides the LC−MS conditions, the injected sample amount can have a significant effect on metabolite detection. Since LC−UV was used to measure the total sample amount, we could readily control the injection so that an optimal amount was used to maximize the number of peak pairs detectable by LC−MS. Figure 2B shows the number of peak pairs detected as a function of the injection volume. There are two sets of peak pair numbers shown in Figure 2B. The first set was generated by counting the level 1 peak pairs, while the second set was from the number of peak pairs detected after zero-filling of the peak pairs detected in all 21 runs. Figure 2B shows that, as the injection volume or amount increases, the peak pair number increases and then levels off after 16 μL. Using all the data sets for zero-filling, a 12 μL injection appears to reach the maximum. On the basis of these results, we concluded that the optimal injection amount for our LC−MS setup was 12 μL of the 4.96 mM fecal solution (i.e., ∼60 nmol). It should be noted that overinjecting samples to LC−MS could cause sample carry-over problem. If happened, more than one washing would be required which increases the overall LC−MS analysis time. In our work, we found that injection of 60 nmol did not cause carry-over problem.

Quantitative Fecal Metabolome Profiling. To demonstrate the performance of the isotope labeling LC−MS platform for fecal metabolome profiling, we analyzed the metabolome differences of the samples collected from three families. In total, 81 ^{12}C-/^{13}C-mixtures were produced from triplicate experiments of 27 samples by using the workflow shown in Figure 1. These mixtures were individually analyzed by LC−MS in triplicate runs. Thus, a total of 243 LC−MS runs were carried out. The MS results were processed using IsoMS[33] to generate a list of peak pairs along with peak ratio and retention time information. Supplemental Table T2 in the Supporting Information shows the final quantitative results of the 243-run data set, which could be exported into a statistics tool for further analysis.

Figure 3A shows the PCA plot of the 243-run data set. Even in this PCA plot, some separation of the three families are visible. Figure 3B shows the OPLS-DA plot of the data according to three family groups; they are clearly separated (R^2X = 0.0681, R^2Y = 0.982, and Q^2 = 0.968; R^2 describes how well the model fits the data and Q^2 estimates predictive ability of the model based on cross-validation results). Figure 3C shows the 3D OPLS-DA plot to examine the individual data points. The person-to-person separation of metabolomic data points within a family is not significant compared to the family to-family separation. Figure 4 shows the 2D and 3D OPLS-DA plots based on the classification of father, mother, and infant for the three families. Although there is a clear separation among the three groups (R^2X = 0.0735, R^2Y = 0.980, and Q^2 = 0.882), the separation is not as good as the family separation, as indicated by the reduced Q^2 value. The 3D plot clearly shows person-to-person separation within a family.

dx.doi.org/10.1021/ac503619q | Anal. Chem. 2015, 87, 829−836

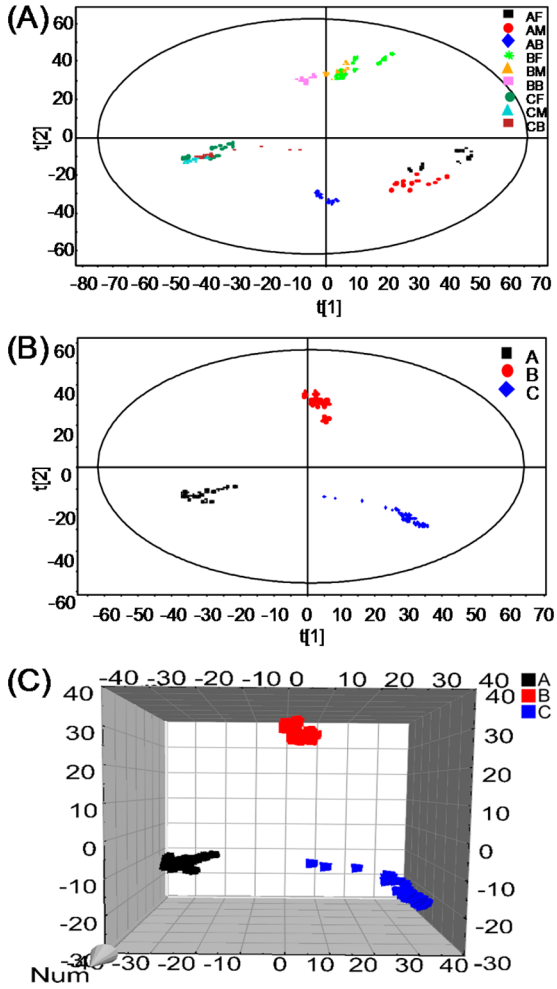


Figure 3. (A) PCA plot of all the data obtained from the 234 LC–MS runs. The data points of individuals are color-coded to show the data distribution. (B) Two-dimensional and (C) three-dimensional OPLS-DA score plots of three family groups.

The day-to-day separation likely caused by diet effect is more visible for individuals within a family, as it is shown in Figure 5. For example, the PCA plot of family B (Figure 5A) shows three separated clusters for three different days of fecal samples from the mother. For family C (Figure 5C), the three clusters of the mother overlap with those of the father to some extent. For the infant within each family, the day-to-day separation is not clearly visible, which can be attributed to the fact that the day-to-day diet for the infant was not varying greatly (e.g., only human milk was taken). Within a family, the father, mother, and infant metabolome data can be readily separated, as shown in the OPLS-DA plots in Supplemental Figure S3 in the Supporting Information.

The above results based on profiling the amine and phenol submetabolomes suggest that diet might affect the metabolome, but the variations among families or even individuals were greater than day-to-day variations. To generalize this finding, we will need to increase the sample size and take fecal samples from different cohorts of individuals prescribed to different diets as well as profile other groups of submetabolomes. Nevertheless, our preliminary results already indicate that the CIL LC–MS method described in this work can provide high precision, as evident for the tight clustering of replicate data points shown in the PCA and OPLS-DA plots (Figures 3–5),

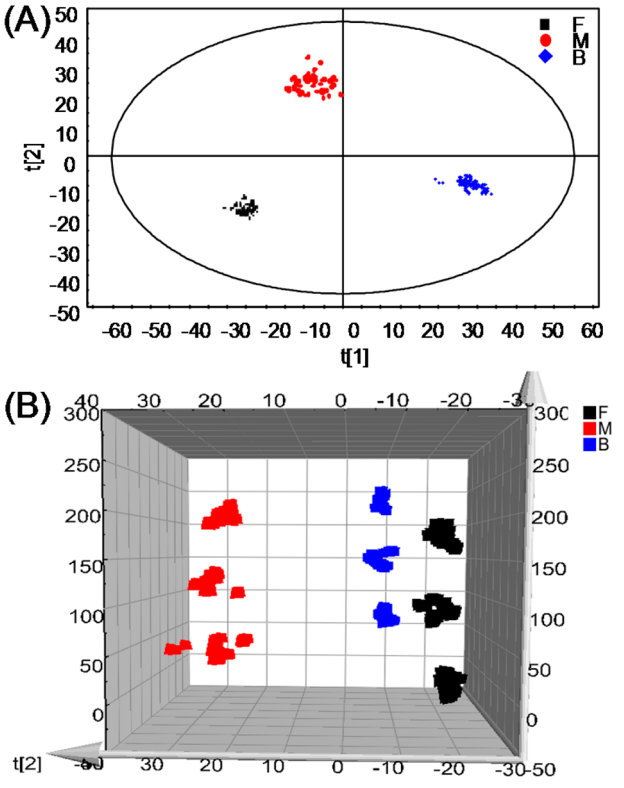


Figure 4. (A) Two-dimensional and (B) three-dimensional OPLS-DA score plots of fathers, mothers, and infants.

to reveal metabolome differences of individuals or different groups. To this end, we have determined the top 20 significant metabolites that provide the binary separation of the fathers, mothers, and infants (Supplemental Table T3 in the Supporting Information) and three families (Supplemental Table T4 in the Supporting Information).

Human Fecal Submetabolome. As Supplemental Table T2 in the Supporting Information shows, we detected an average of 1785 peak pairs or putative metabolites per sample. From the 243 runs, we detected a total of 6200 unique peak pairs. We have examined the distributions of the peak pair numbers detected according to family, father, mother, and infant (see Supplemental Figure S4 in the Supporting Information). Out of 2078 peak pairs found in total, 716 pairs are detected commonly from at least two families and 184 pairs are commonly detected from all three families (Supplemental Figure S4A in the Supporting Information). The peak pair number distributions for the group of fathers (Supplemental Figure S4B in the Supporting Information), mothers (Supplemental Figure S4C in the Supporting Information), or infants (Supplemental Figure S4D in the Supporting Information) are similar; more unique peak pairs are found from an individual than the common peak pairs.

These results indicate that the human fecal metabolome contains a great number of metabolites and the metabolome composition can vary greatly from one individual to another. Fortunately, there are still many common metabolites detectable from all these samples. If we only compare the common peak pairs found from the fecal samples of fathers (317 pairs), mothers (342 pairs), and infants (316 pairs), a total of 469 peak pairs are detected (see Supplemental Figure S5E in the Supporting Information). Among them, 184 (39%) pairs are commonly detected. In applying human fecal metabolomics

dx.doi.org/10.1021/ac503619q | *Anal. Chem.* 2015, 87, 829–836

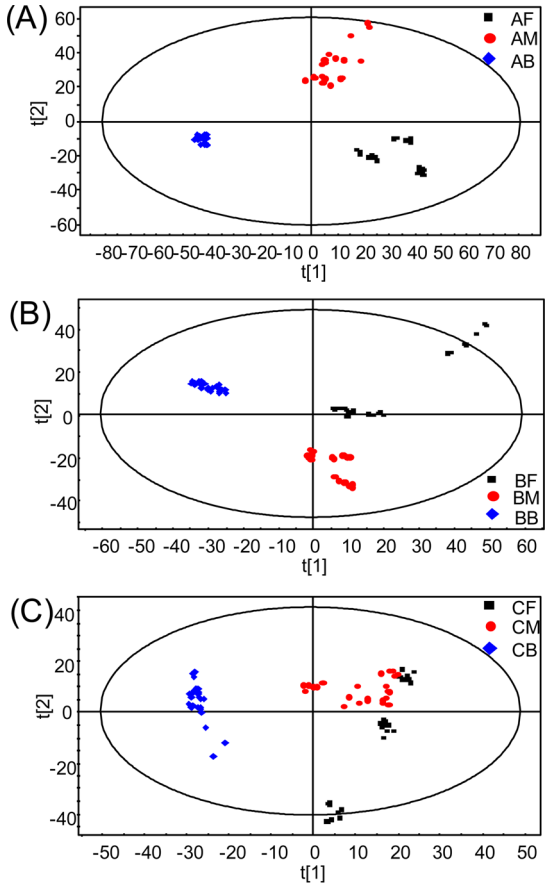


Figure 5. PCA plots of the metabolome data from (A) family A, (B) family B, and (C) family C.

for discovering metabolite biomarkers of diseases, one would hope that there will be specific metabolites consistently detected with significantly higher or lower concentrations than the normal controls. Considering CIL LC−MS can provide high technical reproducibility and high metabolite detectability, we feel that this method is well positioned for disease biomarker discovery work. We note that while untargeted approach is applicable to more chemical groups, it is limited in metabolome coverage for each group. In contrast, the group-based profiling approach using CIL LC−MS can provide much higher coverage of the targeted submetabolome. With the development of other labeling methods in the near future, profiling more group-based submetabolomes will become possible. At this stage, it is perhaps worth pursuing both the untargeted approach and targeted submetabolome approach for disease biomarker study.

Finally, metabolite identification was carried out for all the peak pairs found in this study. At first, the measured mass and retention time of each unique peak pair was searched against a dansyl standard compound library of 262 compounds. We matched 67 metabolites (see Supplemental Table T5 in the Supporting Information for the list) which can be considered to be identified with very high confidence. This list shows the great diversity of metabolites found in fecal samples. We then searched the peak pairs using MyCompoundID against the HMDB and EML compound libraries. On the basis of the mass match, 581 matched to the metabolites in HMDB (Supplemental Table T6 in the Supporting Information) and 3197 matched to the predicted metabolites with one-reaction in EML (Supplemental Table T7 in the Supporting Information) for a total of 3778 matches, representing 60.9% of the total number of peak pairs detected (6200). Although in many cases we could not narrow the match list down to one metabolite candidate, these matches suggest that they are likely from real metabolites not random noises. We note that the use of TOF-MS, instead of tandem MS, does not provide the opportunity for generating MS/MS spectra for potential metabolite identification, and thus metabolite identification in TOF-MS can only be done using a standard library search based on accurate mass and retention time information.

The above results indicate that the human fecal submetabolome of amines and phenols is very complex. The workflow shown in Figure 1 should be, in principal, applicable to profile other types of submetabolomes such as those containing acid groups.[36] Development of labeling reagents targeting other functional groups such as aldehyde and ketone, including various sugars, is currently underway and will be reported in the future.

CONCLUSIONS

We have developed a high-performance chemical isotope labeling LC−MS method for human fecal metabolome profiling. From 243 LC−MS runs of dansyl labeled fecal samples collected from three families, a total of 6200 peak pairs or putative metabolites were detected; 60.9% of them matched to the metabolites in the HMDB and EML metabolite libraries. The metabolome profiles generated were shown to be useful for separating different groups of fecal samples. While the results shown in this work already represent the most comprehensive profile of the amine- and phenol-submetabolome in human fecal samples, future work will be directed toward the analysis of other groups of the fecal metabolome. In addition, the applications of this work for biomarker discovery of diseases and studying the interrelations of the gut metabobiome and human health will be carried out.

ASSOCIATED CONTENT

Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: Liang.Li@ualberta.ca.
*E-mail: ljli@zju.edu.cn.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This collaborative research was in part supported by a visiting professorship to L. Li by Zhejiang University K. P. Chao's Hi-Tech Foundation for Scholars and Scientists. Additional support was provided by the Natural Sciences and Engineering Research Council of Canada, Canadian Institutes of Health Research, Canada Research Chairs, Genome Canada and Alberta Innovates (to Liang Li), the Independent Foundation of the State Key Laboratory for Diagnosis and Treatment of Infectious Diseases and National Science and Technology Major Project of China (Grant 2013ZX10004101-007) and Zhejiang Public Welfare Project, China (Grant 2014C37040) (to Lanjuan Li), and the Open Foundation of the State Key

Laboratory for Diagnosis and Treatment of Infectious Diseases (Grant 2014KF04) (to Nan Wang) and Investigator Project Grant (2010ZZ01) (to Liang Li).

REFERENCES

(1) Ding, T.; Schloss, P. D. *Nature* **2014**, *509*, 357.
(2) Costello, E. K.; Stagaman, K.; Dethlefsen, L.; Bohannan, B. J. M.; Relman, D. A. *Science* **2012**, *336*, 1255−1262.
(3) Kuczynski, J.; Lauber, C. L.; Walters, W. A.; Parfrey, L. W.; Clemente, J. C.; Gevers, D.; Knight, R. *Nat. Rev. Genet.* **2012**, *13*, 47−58.
(4) Blaser, M.; Bork, P.; Fraser, C.; Knight, R.; Wang, J. *Nat. Rev. Microbiol.* **2013**, *11*, 213−217.
(5) Turnbaugh, P. J.; Ley, R. E.; Hamady, M.; Fraser-Liggett, C. M.; Knight, R.; Gordon, J. I. *Nature* **2007**, *449*, 804−810.
(6) Huttenhower, C.; Gevers, D.; Knight, R.; Abubucker, S.; Badger, J. H.; Human Microbiome Project Consortium; et al. *Nature* **2012**, *486*, 207−214.
(7) Li, M.; Wang, B. H.; Zhang, M. H.; Rantalainen, M.; Wang, S. Y.; Zhou, H. K.; Zhang, Y.; Shen, J.; Pang, X. Y.; Zhang, M. L.; Wei, H.; Chen, Y.; Lu, H. F.; Zuo, J.; Su, M. M.; Qiu, Y. P.; Jia, W.; Xiao, C. N.; Smith, L. M.; Yang, S. L.; Holmes, E.; Tang, H. R.; Zhao, G. P.; Nicholson, J. K.; Li, L. J.; Zhao, L. P. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2117−2122.
(8) Claesson, M. J.; Cusack, S.; O'Sullivan, O.; Greene-Diniz, R.; de Weerd, H.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 4586−4591.
(9) Segata, N.; Haake, S. K.; Mannon, P.; Lemon, K. P.; Waldron, L.; Gevers, D.; Huttenhower, C.; Izard, J. *Genome Biol.* **2012**, *13*, 18.
(10) Jacobs, D. M.; Gaudier, E.; van Duynhoven, J.; Vaughan, E. E. *Curr. Drug Metab.* **2009**, *10*, 41−54.
(11) Marcobal, A.; Kashyap, P. C.; Nelson, T. A.; Aronov, P. A.; Donia, M. S.; Spormann, A.; Fischbach, M. A.; Sonnenburg, J. L. *ISME J.* **2013**, *7*, 1933−1943.
(12) Saric, J.; Wang, Y.; Li, J.; Coen, M.; Utzinger, J.; Marchesi, J. R.; Keiser, J.; Veselkov, K.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. *J. Proteome Res.* **2008**, *7*, 352−360.
(13) Jansson, J.; Willing, B.; Lucio, M.; Fekete, A.; Dicksved, J.; Halfvarson, J.; Tysk, C.; Schmitt-Kopplin, P. *PLoS One* **2009**, *4*, 10.
(14) Chow, J.; Panasevich, M. R.; Alexander, D.; Boler, B. M. V.; Serao, M. C. R.; Faber, T. A.; Bauer, L. L.; Fahey, G. C. *J. Proteome Res.* **2014**, *13*, 2534−2542.
(15) Goedert, J. J.; Sampson, J. N.; Moore, S. C.; Xiao, Q.; Xiong, X.; Hayes, R. B.; Ahn, J.; Shi, J.; Sinha, R. *Carcinogenesis* **2014**, *35*, 2089−2096.
(16) Jump, R. L. P.; Polinkovsky, A.; Hurless, K.; Sitzlar, B.; Eckart, K.; Tomas, M.; Deshpande, A.; Nerandzic, M. M.; Donskey, C. J. *PLoS One* **2014**, *9*, e101267.
(17) Le Gall, G.; Noor, S. O.; Ridgway, K.; Scovell, L.; Jamieson, C.; Johnson, I. T.; Colquhoun, I. J.; Kemsley, E. K.; Narbad, A. *J. Proteome Res.* **2011**, *10*, 4208−4218.
(18) Cao, H. C.; Huang, H. J.; Xu, W.; Chen, D. Y.; Yu, J.; Li, J.; Li, L. J. *Anal. Chim. Acta* **2011**, *691*, 68−75.
(19) Zhao, Y. Y.; Zhang, L.; Long, F. Y.; Cheng, X. L.; Bai, X.; Wei, F.; Lin, R. C. *Chem.-Biol. Interact.* **2013**, *201*, 31−38.
(20) Girlanda, R.; Cheema, A. K.; Kaur, P.; Kwon, Y.; Li, A.; Guerra, J.; Matsumoto, C. S.; Zasloff, M.; Fishbein, T. M. *Am. J. Transplant.* **2012**, *12*, S18−S26.
(21) Russell, W. R.; Gratz, S. W.; Duncan, S. H.; Holtrop, G.; Ince, J.; Scobbie, L.; Duncan, G.; Johnstone, A. M.; Lobley, G. E.; Wallace, R. J.; Duthie, G. G.; Flint, H. J. *Am. J. Clin. Nutr.* **2011**, *93*, 1062−1072.
(22) Lampe, J. W.; Navarro, S. L.; Hullar, M. A. J.; Shojaie, A. *Proc. Nutr. Soc.* **2013**, *72*, 207−218.
(23) Calvani, R.; Brasili, E.; Pratico, G.; Capuani, G.; Tomassini, A.; Marini, F.; Sciubba, F.; Finamore, A.; Roselli, M.; Marzetti, E.; Miccheli, A. *Exp. Gerontol.* **2014**, *49*, 5−11.
(24) Gao, X. F.; Pujos-Guillot, E.; Martin, J. F.; Galan, P.; Juste, C.; Jia, W.; Sebedio, J. L. *Anal. Biochem.* **2009**, *393*, 163−175.
(25) Gao, X. F.; Pujos-Guillot, E.; Sebedio, J. L. *Anal. Chem.* **2010**, *82*, 6447−6456.
(26) Wu, J. F.; An, Y. P.; Yao, J. W.; Wang, Y. L.; Tang, H. R. *Analyst* **2010**, *135*, 1023−1030.
(27) Guo, K.; Li, L. *Anal. Chem.* **2009**, *81*, 3919−3932.
(28) Guo, K.; Li, L. *Anal. Chem.* **2011**, *82*, 8789−8793.
(29) Zhou, R.; Li, L. *J. Proteomics* **2014**, in press.
(30) Peng, J.; Chen, Y. T.; Chen, C. L.; Li, L. *Anal. Chem.* **2014**, *86*, 6540−6547.
(31) Zhou, R. K.; Guo, K.; Li, L. *Anal. Chem.* **2014**, *85*, 11532−11539.
(32) Wu, Y.; Li, L. *Anal. Chem.* **2012**, *84*, 10723−10731.
(33) Zhou, R.; Tseng, C. L.; Huan, T.; Li, L. *Anal. Chem.* **2014**, *86*, 4675−4679.
(34) Huan, T.; Li, L. *Anal. Chem.* **2014**, submitted.
(35) Li, L.; Li, R. H.; Zhou, J. J.; Zuniga, A.; Stanislaus, A. E.; Wu, Y. M.; Huan, T.; Zheng, J. M.; Shi, Y.; Wishart, D. S.; Lin, G. H. *Anal. Chem.* **2013**, *85*, 3401−3408.
(36) Guo, K.; Li, L. *Anal. Chem.* **2010**, *82*, 8789−8793.