Spatial, temporal, and inter-subject variation of the metabolome along the human upper intestinal tract

Jacob Folz
University of California, Davis
Rebecca Culver
Stanford University
Juan Morales
UC Davis
Jessica Grembi
Stanford University https://orcid.org/0000-0001-6142-4913
George Triadafilopoulos
Stanford University
David Relman
Stanford University https://orcid.org/0000-0001-8331-1354
KC Huang
Stanford University
Dari Shalon
Envivo Bio
Oliver Fiehn (ofiehn@ucdavis.edu)
University of California, Davis https://orcid.org/0000-0002-6261-8928

Article

Keywords: small intestine, gut microbiome, metabolomics, microbial metabolism, sulfonolipids

Posted Date: October 6th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2099937/v1

Abstract

Most utilization of human diets occurs in the small intestine, which remains largely unstudied. Here, we used a novel non-invasive, ingestible sampling device to probe the spatiotemporal variation of upper intestinal luminal contents during routine daily digestion in 15 healthy subjects. We analyzed 274 intestinal samples and 60 corresponding stool homogenates by combining five metabolomics assays and 16S rRNA sequencing. We identified 1,909 metabolites, including sulfonolipids and novel bile acids. Stool and intestinal metabolomes differed dramatically. Food metabolites displayed known differences and trends in dietary biomarkers, unexpected increases in dicarboxylic acids along the intestinal tract, and a positive association between luminal keto acids and fruit intake. Diet-derived and microbially linked metabolites accounted for the largest inter-subject differences. Interestingly, subjects exhibited large variation in levels of bioactive fatty acid esters of hydroxy fatty acids (FAHFAs) and sulfonolipids. Two subjects who had taken antibiotics within 6 months prior to sampling showed markedly different patterns in these and other microbially related metabolites; from this variation, we identified Blautia species as most likely to be involved in FAHFA metabolism. Thus, in vivo sampling of the human small intestine under physiologic conditions can reveal links between diet, host and microbial metabolism.

Introduction

The human small intestine is an approximately 6 meter-long tube between the stomach and the colon and is the primary site of protein, fat, carbohydrate, and other nutrient absorption[1]. Malabsorption of dietary components in the small intestine can lead to malnutrition, which is associated with mortality, particularly in older adults[2]. Small intestinal diseases such as gluten-sensitive enteropathy and Crohn's disease (CD) can lead to malabsorption and malnutrition[3]. Their symptoms are non-specific and highly variable and diagnosis requires invasive endoscopic procedures[4,5]. Non-invasive disease screening and better understanding of small intestinal tract diseases will improve human health and quality of life and will open doors for therapeutic interventions by, for example, directly targeting specific microbial compositions. However, progress in this area has been hindered by difficulty in intestinal sampling[6]. Here, we present results from a study performed using a newly-developed ingestible device that non-invasively collects samples of luminal contents from the small intestine and ascending colon. This technology should facilitate intestinal disease diagnosis, monitoring of nutrient absorption across the small intestine, and improve understanding of the complex human gut ecosystem.

Metabolites in the small intestine originate from host secretion, anything we ingest (the exposome)[7], and microbial transformations of the above. Humans secrete complex mixtures of chemicals and enzymes into the gastrointestinal tract via saliva[8], gastric secretions[9], bile[10], pancreatic fluid[11], and epithelial mucus[12]. Commensal gut microbes extensively modify metabolites causing an immense expansion in chemical diversity. Non-enzymatic chemical reactions further increase this chemical space[13].

Metabolomics analysis using mass spectrometry and comprehensive databases is most suitable to chemically analyze such complex biological mixtures[14,15].

Here, we showcase the utility of analyzing intestinal tract samples from humans under a normal dietary routine with respect to meal size, frequency, and composition. We demonstrate that these data can be used to reconstruct transformations of protein, carbohydrate, and fat digestion. We further provide evidence that the sampling devices targeted the intended locations within the gut and offer insights into specific microbes that alter the gut metabolome.

Results

Sampling and metabolomic analyses

We aimed to comprehensively study metabolomic differences among luminal samples from the upper intestinal tract of 15 healthy subjects to better understand the extent of spatial and temporal variation and to gauge the prospects of integrating metabolome and microbiome data. Volunteers swallowed sets of 4 sampling devices per sampling timepoint. These ingestible sampling devices were comprised of a collapsed collection bladder capped by a one-way valve in a capsule treated with pH-sensitive coatings. The four types of capsules differed only in their enteric coating which dissolved at pH 5.5 (capsule 1), pH 6 (capsule 2), and pH 7.5 (capsules 3 and 4) (**Figure 1A**). The thickness and pH-responsiveness of the coating enabled sampling at specific locations of the intestinal tract after entry into the duodenum. The devices did not contain any electronics beyond a passive radio frequency identification chip for tracking purposes. Once the coatings dissolved, an elastic collection bladder expanded and collected up to 400 µL of luminal contents through vacuum suction. The one-way valve prevented loss of sample and contamination from downstream fluids. Stool samples were frozen at -20 °C and all capsules were recovered from the stool prior to analysis. Liquid contents were retrieved from capsules using hypodermic needles. Aliquots of the raw sample were used for 16S ribosomal RNA microbiome analyses and the supernatants from centrifugated samples were used for metabolomic studies.

The measured pH of the luminal contents for device types 1 through 4 were consistent with the expected pH gradient across the intestinal tract[16,17], covering the duodenum, jejunum, ileum, and ascending colon (**Figure 1A**). The pH in type 1 and 2 devices was significantly different from type 3 and 4 devices (Figure S1B; Wilcoxon 2-way rank-sum test, p=2.4×10^{-14}), whereas pH was not significantly different between type 1 and type 2 devices, or between type 3 and type 4 devices (Figure S1B). We therefore associated type 1/2 and 3/4 devices with proximal and distal regions of the upper intestinal tract, respectively.

We used five mass spectrometry assays to analyze the luminal contents within these capsule devices and the associated stool samples. By matching chromatographic retention times, accurate precursor masses, and mass spectrometric fragmentation (MS/MS) to MassBank.us public and NIST20 licensed libraries, we annotated 1,909 chemicals from gut luminal and stool contents at MSI confidence levels 1-3 (Table S1)[18], including 155 internal standards used for quality control and quantification purposes. Additionally, >12,000 unknown chromatographic features were reliably detected above the level of method blanks (Table S2). Using ClassyFire software[19], structurally annotated metabolites fell into 61 chemical subclasses (Table S1). Two untargeted high-resolution liquid chromatography (LC) MS/MS assays focusing on hydrophilic and lipophilic metabolites yielded most of the annotated compounds, with 1,608 identifications. Untargeted gas chromatography-MS added 119 primary metabolites, supplemented by targeting 6 short chain fatty acids (SCFA) and a targeted LC-MS/MS assay for 17 bile acids (**Figure 1B**). Quality control analysis of total metabolic variance revealed separation of stool and intestinal samples, with strong clustering of pooled quality control samples (Figure S1A).

Spatial variation across the upper intestine

Metabolome results revealed striking differences between stool and intestinal samples (Figure S2) and among the intestinal tract samples (Figure S3). To uncover spatial differences across the intestine, we applied linear mixed effect models (LMM) that accounted for sampling location (proximal or distal) as well as other variables (Table S3, S4). Specifically, we studied the 1,182 most prevalent metabolites that were detected in >50% of device samples. Of these, 630 (54%) were significantly different in the proximal compared to distal upper intestine (FDR $p < 0.05$; LMM) (**Figure 2**, Table S4), with 473 metabolites at higher levels in the proximal compared to distal upper intestine and 157 compounds at lower levels in the proximal compared to distal upper intestine (**Figure 2**). Known microbially generated chemicals including SCFAs[20,21], secondary bile acids[22], and some microbially conjugated bile acids[23,24], increased from the proximal to distal upper intestine (**Table 1, Figure 2**). Of the 11 detected acetylated amino acids, 7 increased from the proximal to distal upper intestine (raw $p < 0.05$; LMM) (**Table 1, Figure 2**). We also examined the 12,346 chemically unannotated metabolite signals, restricting our attention to 9,317 signals that were detected in >50% of intestinal samples. Overall, 3,594 (38%) features were significantly different between the proximal and distal upper intestine, with 1,937 features at higher levels in the proximal compared to distal upper intestine and 1,657 features at lower levels in the proximal compared to distal upper intestine (FDR $p < 0.05$; LMM) (Figure S4).

To interrogate general metabolic differences between locations, we used chemical enrichment statistics. Di- and tri-peptides were among the most significantly decreased classes from the proximal to distal upper intestine (**Table 1,** Figure S5). Of the 333 di- and tri-peptides measured, 262 significantly decreased in abundance from the proximal to distal upper intestine (raw $p < 0.05$; LMM) (**Table 1**). A range of other

compound classes also exhibited significantly higher levels in proximal intestinal tract samples compared with distal samples, including sugars, sugar alcohols, nucleosides, carnitines, and ceramides (**Table 1**, Figure S5). Several other compound classes exhibited significantly lower levels in proximal compared to distal intestinal tract samples, including dicarboxylic acids and phenolics (**Table 1**, **Figure 2**). In fact, 3 of the top 6 overall most significantly increased metabolites between the proximal and distal upper intestine were dicarboxylic acids (hexadecanedioic acid, tetradecanedioic acid, and octadecanedoic acid) (**Figure 2**).

Food-specific metabolites

Next, we used LMM to test for associations of food intake logs recorded by the subjects to levels of intestinal tract metabolites. We tested for consumption of fruit, alcohol, dessert, animal protein, vegetables, grains, coffee/tea, and dairy food types ingested 6 hours before swallowing capsule devices (Table S3). After correcting for multiple-hypothesis testing, some food types had no significantly associated metabolites, unsurprisingly due to the small sample size for some food types and strong FDR correction accounting for tests of 1,182 metabolites (**Table 3, Figure 3,** Table S4). Nonetheless, our analysis revealed several strong associations, some of which validated previous biomarkers and others that had not previously been identified.

Fruit-associated metabolites

Using effect size differences of ±0.2 and raw $p < 0.05$, fruit consumption was significantly associated with 20 compounds at increased concentration and 17 metabolites at decreased concentration (**Figure 3A**). Some metabolites were directly linked to fruit consumption based on having strict FDR-corrected $p < 0.05$ (**Table 2**, **Figure 3A**), including N-methylproline and stachydrine, both of which were previously reported as fruit consumption biomarkers for blood plasma in non-controlled dietary studies[25]. Betonicine, a known component of fruit juice[26], also increased after fruit consumption at raw $p < 0.05$ (**Figure 3A, Table 2**) but did not achieve the FDR-significance threshold. Similarly, three keto acids (4-methyl-2-oxovaleric acid, ketoisovaleric acid, and 3-methyl-2-oxovaleric acid) also significantly increased in response to fruit intake at raw $p < 0.05$ (**Figure 3A**, **Table 3**). Since metabolites are not independent from each other, rather they are linked via food compositions and microbial and enzymatic pathways, we used ChemRICH chemical set enrichment statistics to identify significantly altered clusters of metabolites (**Figure 3C**). This strategy revealed keto acids as the chemical class with the most significant response to fruit (**Figure 3A,C**), representing the first report of keto acids as a fruit biomarker in the human gut. Importantly, ChemRICH also revealed that typical fruit ingredients like phenylacetates and phenolic natural products were positively associated with fruit intake (**Figure 3C**).

Alcohol-associated metabolites

Alcohol consumption was most significantly associated with ethyl sulfate (FDR $p < 0.05$), a known plasma biomarker of alcohol consumption (**Table 2**)[27]. Stachydrine was linked with both fruit and alcohol consumption (FDR $p < 0.05$). Trp-Lys was significantly decreased with alcohol consumption after FDR correction (**Figure 3B**). In total, 40 di- and tri-peptides decreased with alcohol consumption (raw $p < 0.05$) with a ChemRICH cluster $p = 8.8\times10^{-18}$ (Table S4, Figure S6, S7).

Dessert-associated metabolites

"Dessert" was defined as consumption of high fat/high sugar foods, such as soda, cake, and ice cream. Two substituted benzoic acids, 3-hydroxy-4-methoxybenzoic acid and 3,4-dihydroxybenzoic acid, were associated with dessert (raw $p < 0.05$) (**Table 2, Figure 3D**). These compounds are metabolic intermediates in the breakdown of vanillin and isovanillin[28]. Neochlorogenic acid was also significantly associated with dessert (raw $p < 0.05$). Neochlorogenic acid is present in a variety of fruits and berries[29] including cherries[30] and peaches[31]. Other food types that were included in the mixed effect model also had significantly associated metabolites (Table S4, Figure S6, S7).

Caffeine metabolism

Caffeine was detected in the vast majority of samples (**Table S1**). Surprisingly, caffeine was not significantly associated with coffee or tea consumption during the experimental timeframe (FDR $p = 0.87$) (Table S4), most likely due to the fast absorption of caffeine after ingestion. However, caffeine metabolic pathways were readily discerned through Spearman-rank correlation analysis. The six metabolites that were most strongly correlated at FDR $p < 10^{-13}$ to caffeine were known caffeine catabolites, including theophylline and theobromine (**Figure 4A,B**). Hence, upper intestinal tract metabolite correlations may enable reconstruction of microbial and enzymatic pathways of food degradation, especially because the detected levels of food compounds, like caffeine products, can span four orders of magnitude.

Temporal variation across upper intestinal samples

To investigate whether sampling time or sampling region had a larger impact on upper intestinal metabolites, we used analysis of variance to calculate the number of metabolites that significantly differed between the 4 device types by subject, or between the 4 sampling time points (after each meal) by subject. The large differences in metabolite levels between the proximal and distal sampling regions were often superseded by metabolic differences between time points, showing that 12 of 15 subjects had

more statistically different metabolites between meals (time points) than between intestinal regions (capsule types) (**Figure 5A**). A closer inspection of the compound classes that contributed to these differences found that di- and tri- peptides (within the chemical class of carboxylic acids) were the largest chemical class that distinguished between capsule types, representing >70% of all significantly different metabolites in 5 subjects and >40% for another 7 subjects (Figure S8A). For metabolites that differentiated sampling time points, we found sugars (organooxygen compounds) to be enriched in 13 of 15 subjects (Figure S8B). Similarly, more significantly different imidazopyrimidines, indoles, and isoflavanoids were found to distinguish sampling time points than intestinal regions (Figure S8). These classes signify dietary metabolites that were different due to variation between food types ingested during different meals, but were not as useful for differentiating between intestinal regions.

Overall metabolome variation among subjects

Our data showed large inter-subject variation (Figure S8). Using multivariate discriminant analysis (PLS-DA), we identified large differences in the proportion of metabolites that were most important to differentiate between the proximal and distal intestine (capsule types) and between the 15 subjects (**Figure 5B**). The large overall variance among samples obscured clear visualization of PLS-DA based on subjects, capsules, or time points. Nonetheless, the 100 metabolites that contributed most to multivariate discrimination included many diet/exposome-related compounds including caffeine and caffeine breakdown products, acetaminophen, ethyl sulfate, capsaicin, and secoisolariciresinol (Table S5). Using chemical subclass categories[19], carbohydrates and purines were more important for differentiating among subjects than among sampling regions (**Figure 5B**). These data support the hypothesis that diet is a critical factor causing inter-individual differences in intestinal lumen metabolites. Similarly, peptides were more important for differentiating between sampling regions than between subjects. Indeed, no di- or tripeptides were found in the 100 metabolites that contributed most to multivariate discrimination between subjects (Table S5). Interestingly, hierarchical clustering separated stool samples by subject while intestinal samples did not strongly cluster by subject (Figure S9).

Variation in bile pigments among subjects

Although large overall variation obscured direct visualization of inter-subject differences when using all data in PLS-DA projections, specific compounds exhibited very large concentration differences among individuals (**Figure 6**). For example, the human heme-derived bile pigments biliverdin and bilirubin and the microbially produced urobilin and stercobilin varied drastically among capsules for some subjects and were also greatly reduced or absent in specific subjects, such as stercobilin for subjects 10 and 15 (**Figure 6**). Interestingly, these two subjects (#10 and 15) reported use of antibiotics 1-5 months prior to the study. Conversely, subjects 5 and 7 showed high levels of stercobilin but frequent absence of biliverdin and bilirubin. The same subject-specific profiles of bile pigments were also observed in stool samples (Figure

S10). Production of secondary bile acids has been proposed to use a similar enzymatic pathway as stercobilin production[32,33] and the two subjects that had consumed antibiotics also showed reduced levels of deoxycholic acid, a secondary bile acid (**Figure 4**). These data support the hypothesis that metabolic profiles reflect differences in microbial activity for specific pathways, here demonstrated for bile pigment metabolism[34]. We did not identify significant associations of microbial species abundance in 16S rRNA gene quantification to stercobilin levels, likely due to limited statistical power.

Variation in fatty acid esters of hydroxy fatty acids (FAHFAs) among subjects

FAHFAs were first identified 10 years ago and are biologically active and regulate physiology[35,36]. A specific subset of FAHFAs with an alpha lipid linkage are acyl alpha-hydroxyl fatty acid (AAHFAs) and were discovered just 2 years ago[37]. We identified 88 FAHFAs in this study (Table S1). Strikingly, a very large difference among subjects was found for four specific FAHFAs, all of them linkages of a long-chain hydroxyl fatty acid back bone that was esterified with C3- or C4-short chain moieties. Three of these bioactive lipids were esterified at the alpha-position ('AAHFA'; AAHFA 4:0/22:0, AAHFA 3:0/22:0, and AAHFA 3:0/24:0), and one was esterified elsewhere on the backbone (FAHFA 3:0/23:0). Nine subjects frequently exhibited high levels of these FAHFA lipids while 6 subjects produced very little or undetectable amounts (**Figure 6**). The same subject-specific trends were observed in stool samples (Figure S10).

Interestingly, low levels of long-chain or short-chain fatty acid substrates propionic (C3:0) and butyric acid (C4:0) did not explain the differences observed in these short-chain FAHFAs in intestinal or stool samples (**Figure 6**, Figure S11). We therefore investigated whether specific bacteria were associated with the presence or absence of the four FAHFAs of interest. Only one taxa, *Blautia sp,* phylogenetically most closely related to *Blautia obeum,* was significantly associated (FDR $p < 0.0001$) with detection of these FAHFAs. *Blautia obeum* is a known short chain fatty acid producer[38] which initially suggested production of the short chain fatty acyl constituents as a driver of the subject-specific FAHFA production. However, FAHFA 4:0/16:0 did not show the same subject-specific trend, suggesting a high substrate specificity for the enzyme(s) that are likely involved in FAHFA production in *Blautia sp.*

Variation in sulfonolipids among subjects

For the first time, here we report the detection of sulfonolipids in human samples. These lipids were only recently added to lipidomic libraries[39], leading to the discovery that they are microbially produced in mouse intestinal tracts and linked with inflammatory phenotypes[40,41]. In intestinal device samples, we detected sulfonolipids with strong inter-subject trends regardless of sampling time point (meals), suggesting that sulfonolipids were microbially produced in some subjects, but not in others. The 2

subjects who had previously received antibiotics showed much lower levels of sulfonolipids than all other subjects (**Figure 6**), suggesting that the antibiotic treatment may have eliminated the microbial producers. Interestingly, sulfonolipids were also absent in specific samples of other subjects, prompting us to examine associations of sulfonolipids with bacterial taxa. Sixteen taxa were significantly enriched in sulfonolipid-containing samples, including 3 members of the Desulfovibironaceae family (Table S6). Also enriched were 4 members of Bacteroidetes, a phylum with known sulfonolipid producers, including an *Alistipes* species that was previously associated with sulfonolipid production in mice[41].

Intestinal samples compared to stool samples

The chemical profiles of intestinal samples differed substantially from those of stool (Figure S2). Thirty-one metabolites were >100 times more abundant on average in the intestine compared with stool. These metabolites consisted of glycinated lipids, sugars, plant natural products, carnitines, microbially conjugated bile acids, and S-succinylcysteine (Table S7). Peptides were also generally at much lower levels in stool samples compared with intestinal samples, especially when compared to the proximal intestine (Figure S2). Microbially conjugated bile acids were dramatically higher in the upper intestine compared to stool, despite the much lower microbial density in the small intestine compared to the colon. These data confirm the involvement of microbially conjugated bile acids in enterohepatic circulation. We also identified >100 metabolites that were >100 times more abundant in stool compared with intestinal samples (Table S7); these metabolites were mostly polar lipids such as PEs, PIs, and PGs, as well as specific FAHFAs. The high abundance of membrane lipids in stool samples is likely due to the high amount of bacterial cell material compared to luminal samples from the upper intestine.

Discussion

Spatial and temporal variation in the intestinal tract

Here, we provide the first report of metabolome differences in the upper intestinal tract in healthy human subjects using a non-invasive, ingestible sampling device. The results presented here open the door for future detailed *in vivo* studies on digestion and intestinal diseases. As expected, the metabolome of stool was highly distinct from that of the intestine. Thus, stool cannot serve as a surrogate for the gut intestinal tract, rather only for colonic contents (at best). Even within the intestinal tract, >50% of annotated metabolites exhibited significantly different levels between proximal and distal locations. These spatial differences in the intestine reflect classic food digestion and absorption[42] that we observed for di- and tripeptides[43] and acylcarnitines[44,45], or ceramides that are hydrolyzed to sphingosine and free fatty acids prior to intestinal uptake[46]. By contrast, short chain fatty acids exhibited increased levels in distal regions, likely due to their production by microbes[20,21]. Acetylated amino acids, which have been associated with Crohn's Disease[47], also exhibited higher levels in the distal compared to proximal upper intestine, possibly

due to slower absorption of acetylated compared with non-acetylated amino acids[48,49]. Bile acids are transformed extensively by microbes, and levels of secondary bile acids increased along the intestine[17].

We also found that dicarboxylic acids increased in concentration from the proximal to distal upper intestine. Dicarboxylic acids are generated during catabolism (omega-oxidation) of fatty acids, which occurs in human cells[50], plants[51], and microbes[52,53]. The lead compound hexadecanedioic acid was most strongly correlated with other dicarboxylic acids, plant metabolites, bile acids, and known microbially produced compounds (Table S5). Epithelial cells contain omega-hydroxylated lipids essential to maintain epithelium barrier function[54] that can be cleaved by lipases to form dicarboxylic acids. We hypothesize that the consistent and significant increase of dicarboxylic acids along the upper intestine is due to catabolism of human epithelial lipids.

We found that dietary metabolites were a leading cause of temporal differences during the 2 days and 4 sampling time points of this study. Subjects were not prescribed specific meals, and hence diet-based variation was expected to differentiate subjects and time points. Future studies with longer sampling and/or controlled diets may elucidate patterns, diurnal cycles, or temporal shifts in microbial populations involved in small intestinal health and disease.

Food-related metabolites

The bioavailability, absorption, and mechanism of action of dietary phenolic metabolites are not fully understood[55-57]. We measured 7 and 28 phenolic metabolites that decreased and increased, respectively, from the proximal to distal upper intestine (Table S4). These trends are likely caused by a combination of factors including intestinal absorption, enzymatic transformation[57,58] such as deglycosylation, and by delayed breakdown of plant cells and cell-wall components by microbial enzymes[59-61]. As an example, we found that the flaxseed-affiliated lignan secoisolariciresinol was most significantly enriched in distal regions compared with proximal samples, likely due to both bioavailability[62] and deglycosylation[63].

Despite the small size of this study with 15 subjects, we were able to validate a range of dietary biomarkers that were previously found in blood and correlated with fruit[25] and alcohol[27] consumption. These examples serve as proof-of-concept, lending credibility to our discovery of keto acids as biomarkers for fruit intake. Keto acids are formed from enzymatic deamination of amino acids, for example by gut bacteria[64]. The decrease in di- and tri-peptides after alcohol consumption suggested a

decrease in total protease activity, possibly due to impaired pancreatic secretion[65,66], rather than direct inhibition because trypsin and chymotrypsin are active even in 20% ethanol solution[67].

Both human[68] and bacterial[69,70] enzymes can transform caffeine into a number of chemical products. Our association of caffeine levels with caffeine catabolites shows how other exposome biochemical pathways can be investigated in the human gut in future targeted studies using this sampling approach. We did not find an association of coffee intake and intestinal caffeine levels, likely because caffeine is absorbed rapidly within 1 h of oral intake, and has a mean half-life of 4.5 h (range of 2.7 to 9.9 h) in the bloodstream[71]. Caffeine is metabolized and excreted through several routes including urine[72] and bile[73]. Bile is the expected origin of caffeine measured in this study, since multiple hours passed between beverage consumption and capsule sampling events.

Inter-subject variation

Subject-specific trends were best explained by differences in the levels of metabolites of plant and microbial origin (**Figure 5B**), including the pepper compound capsaicin, the flaxseed compound secoisolariciresinol, and the microbially produced butyric acid and propionic acid. Other metabolites that showed subject-specific variation, such as N-methylhistamine, phenethylamine, phenylacetaldehyde, and succinic acid, may depend on a combination of human, dietary, or microbial factors.

Interestingly, we identified a subset of microbially linked metabolites that were highly subject-dependent, particularly in two subjects who received oral antibiotics within 5 months prior to sampling. These subjects were characterized by very low levels of stercobilin, deoxycholic acid, a subset of FAHFAs, and sulfonolipids, all of which have been linked to the gut microbiota[22,37,74]. FAHFAs have been strongly linked to the etiology of diabetes[36]. FAHFAs are also found endogenously[36] and in food[75]. Utilizing the varying presence of these metabolites across samples and subjects, we related specific bacterial taxa to metabolites. A group of four FAHFAs were significantly associated with *Blautia sp*. Highly subject-specific FAHFAs were detected, with a combination of very long-chain (≥22 carbons) and short-chain (3-4 carbons) fatty acid (SCFA) components. Other FAHFAs in our study contained two long-chain fatty acyls (16-18 carbons). The only other frequently detected FAHFA with butyric acid and a long-chain fatty acid was AAHFA 16:0/4:0, which displayed a different subjectspecific pattern than the four *Blautia sp*-associated FAHFAs. We therefore propose that *Blautia sp* specifically produces FAHFAs with propionic and butyric acid esters of hydroxylated very long-chain fatty acyls (22:0, 23:0) and not of hydroxyl-forms of the most abundant fatty acids, C16-18.

We also detected sulfonolipids in humans for the first time. Sulfonolipids were associated with 15 microbial taxa, including known sulfonolipid producers. Interestingly, we also identified three bacterial species in the Desulfovibironaceae family, which has been associated with ulcerative colitis[76]. Sulfonolipids have been associated with high-fat diets[41] and have both pro- and anti-inflammatory effects[40,77]. In this study, two subjects lacked the proposed sulfonolipid-producing bacteria, and we did not detect sulfonolipids in their samples.

An important question is the time frame needed for repopulation of subjects with sulfonolipid-, stercobilin- and long-chain AAHFA-producing bacteria after antibiotic treatments. The prevalence and abundance of these bacteria in the human gut, which is impacted by antibiotics, may correlate with the incidence and etiology of diabetes and inflammatory bowel disease. In summary, the use of non-invasive sampling devices, in combination with metabolomics and genomics, has substantial potential to enable more precise intervention and prevention strategies for addressing human disease.

Methods

Study design

The study was approved by the WIRB-Copernicus Group IRB (study #1186513) and written informed consent was obtained from each subject. Healthy volunteers (**Table 3**) were selected to exclude participants suffering from clinically detectable gastrointestinal conditions or diseases that would potentially interfere with data acquisition and interpretation. Recruitment and sample collection was carried out between March of 2019 and October of 2020 in the San Francisco Bay Area.

Patient and Public Involvement

Subjects and the public did not contribute to the research question or study design**.** Healthy volunteers were recruited from a routine clinical practice setting without any public advertisement. Subjects were aware of the research questions and understood the burden and time required to participate in the study prior to their enrollment.

Sample collection

The CapScan sampling devices (Envivo Bio Inc, San Carlos CA) were constructed with a coating designed to dissolve at a specific pH to take advantage of the pH gradient of the human intestine. After the coating dissolved, a compressed elastic bladder expanded to pull in 400 µL of luminal contents through a one-way valve. This valve remained sealed until recovery from stool. The pH coating of each capsule type

dissolved at pH 5.5 (type 1), 6 (type 2), or 7.5 (types 3 and 4). Type 4 also had a time delay to target the distal ileum or ascending colon. Four sampling capsules were swallowed 3 hours after lunch or dinner across 2 days (Figure 1A). Subjects were instructed to maintain their normal diet, record the time of any food or drink consumed over the testing period, and to not consume caffeinated beverages after lunch on sampling days. Detailed guidelines are provided in Supplemental Material. Stool was collected and immediately frozen at -20 °C until stool was thawed and filled capsule devices were retrieved. Liquid sample was removed from each bladder using a hypodermic needle. An aliquot of each sample was used for 16S rRNA gene sequencing while another aliquot was centrifuged at 10,000 rcf for 3 min, and the supernatant was used for metabolomics analysis.

Sample preparation

For all non-targeted analyses, 10 µL of intestinal lumen samples were subjected to a modified biphasic water, methanol, and methyl tert-butyl ether extraction[78] to separate polar and non-polar metabolites. The polar and non-polar phases were divided into multiple aliquots in 96-well plates, dried by rotary vacuum, and frozen until further analysis. Homogenized stool samples were prepared using an analogous extraction procedure with modification for bead-homogenization and extraction in microcentrifuge tubes to account for the solid nature of the sample. Targeted bile acid analysis was performed using aqueous phase of the described biphasic extraction. Targeted SCFA analysis used an acidified water and MTBE extraction followed by MTBSTFA derivatization[79]. For detailed sample preparation methods see Supplemental Material.

Data acquisition and statistics

Untargeted mass spectrometry-based metabolomics was performed using reversed-phase and hydrophilic interaction liquid chromatography coupled to high resolution mass spectrometry (LC-MS/MS) and gas chromatography-time of flight mass spectrometry (GC-MS)[80]. Targeted metabolite analyses were conducted by unit resolution LC-MS/MS and GC/MS[80]. Details are given in Supplemental Material and Table S9.

Microbiota composition was analyzed as described previously[81]. In brief, DNA was extracted using a Microbial DNA extraction kit (Qiagen) and 50 µL from a capsule device, 200 µL of saliva, or 100 mg of stool. 16S rRNA amplicons were generated using Earth Microbiome Project-recommended 515F/806R primer pairs and 5PRIME HotMasterMix (Quantabio 2200410) with the following program in a thermocycler: 94 °C for 3 min, 35 cycles of [94 °C for 45 s, 50 °C for 60 s, and 72 °C for 90 s], followed by 72 °C for 10 min. PCR products were cleaned, quantified, and pooled using the UltraClean 96 PCR

Cleanup kit (Qiagen 12596-4) and Quant-iT dsDNA High Sensitivity Assay kit (Invitrogen Q33120). Samples were sequenced with 250-bp reads on a MiSeq instrument (Illumina).

Sequence data were de-multiplexed using the Illumina bcl2fastq algorithm at the Chan Zuckerberg BioHub Sequencing facility. Subsequent processing was performed using the R statistical computing environment (v. 4.0.3)[82] and DADA2 as previously described using pseudo-pooling[83]. Taxonomy was assigned using the Silva rRNA database v. 132[84]. Spearman rank was used for correlation analyses. Benjamini-Hochberg[85] false-discovery rate calculations were used to correct for multiple testing.

Declarations

Acknowledgements

Jeremiah Wells and Zakery Tippins assisted in mass spectrometry analyses at the West Coast Metabolomics Center. R code from Dr. Dinesh Barupal was used as downloaded from his GitHub site for ChemRICH assessments. Dr. Christopher Brydges consulted on statistical regression analyses.

Contributors

OF, KCH, and DS designed the study. JF acquired and processed metabolome data, assisted by JMM. JF performed statistical analyses and produced Figures and Tables. GT recruited subjects and obtained consents and samples. KCH, RNC, JG, and DAR acquired, processed, and analyzed microbiome data. JF and OF wrote the manuscript with contributions from DS, KCH, GT, and RNC. All authors approved the final version of the manuscript.

Funding

This study was funded by USDA 2021-67017-35783 to OF and salary for JF was provided by grants NIH R01 AT010216 and NIH U19 AG023122. This material is also based upon work supported by the National Science Foundation under Grant No. 1936687 to DS. KCH and DAR are Chan Zuckerberg Biohub Investigators.

Competing interests

DS is an employee and has equity interest in Envivo Bio Inc., a company with interest in the human gastrointestinal tract.

Ethics approval

The study was approved by the WIRB-Copernicus Group IRB (study #1186513) and informed consent was obtained from each subject.

Data availability

Raw mass spectrometry data is available on the Metabolomics Workbench (https://www.metabolomicsworkbench.org/) under studies ST002073, ST002075.

ORCiDs

Jacob Folz https://orcid.org/0000-0001-5763-2963

Kerwyn Casey Huang https://orcid.org/0000-0002-8043-8138

Dari Shalon https://orcid.org/0000-0003-2646-6215

Oliver Fiehn https://orcid.org/0000-0002-6261-8928

Tables

Table 1: The 20 chemical classes that exhibited the most significant differences between the proximal and distal upper intestine. Significance was calculated by ChemRICH chemical enrichment statistics using univariate statistical results from LMM as input.

Chemical class name	size	*p*-value	FDR *p*-value	example compound	# metabolites increased from proximal to distal	# metabolites decreased from proximal to distal
di-peptides	268	<2.2E-20	2.2E-19	Lys-Phe	10	211
tri-peptides	65	<2.2E-20	2.2E-19	Ala-Ala-Ala	1	51
phenolic natural products	107	<2.2E-20	2.2E-19	secoisolariciresinol	28	7
unsaturated ceramides	121	<2.2E-20	2.2E-19	Cer 42:1;3O| Ceramide 20:1;2O/22:0;O	0	33
conjugated bile acids	35	<2.2E-20	2.2E-19	glutamyl-cholic acid	11	14
carnitines	27	<2.2E-20	2.2E-19	oleoyl-carnitine	0	21
dicarboxylic acids	21	<2.2E-20	2.2E-19	hexadecanedioic acid	18	0
Saturated ceramides	72	1.1E-15	9.9E-15	Cer 34:0;3O| Ceramide 18:0;2O/16:0;O	0	20
unsaturated fatty acids	25	1.1E-13	7.7E-13	9-oxo-10,12-octadecadienoic acid	6	11
sugars	68	1.5E-13	9.7E-13	tagatose	8	22
nucleoside related metabolites	62	7.4E-10	3.9E-09	2‘-O-methyluridine	8	19
amino acids	28	3.0E-08	1.3E-07	beta-alanine	8	5
amino acids, sulfur	10	2.3E-07	9.5E-07	cysteic acid	4	3
saturated fatty acids	15	2.8E-07	1.1E-06	nonadecanoic acid	5	3
phenylacetates	8	3.0E-07	1.1E-06	phenylacetic acid	7	0
unsaturated phospha-tidyl-ethanolamines	44	7.3E-06	2.4E-05	LPE-N (FA)36:4| LPE-N (18:2/18:2)	0	7
sugar alcohols	18	8.5E-06	2.6E-05	deoxypentitol	1	7
FAHFAs	88	3.3E-05	9.0E-05	AAHFA 16:0;O| AAHFA 7:0/9:0;O	4	6
acetylated amino acids	11	4.3E-05	1.1E-04	N-acetyl-valine	7	0
SCFAs	8	9.2E-	2.2E-	acetic acid	5	0

		04	03			

Table 2: Selected metabolites (of 1,182) associated with alcohol, dessert, and fruit consumption. LMM coefficient, raw *p*-value, and Benjamini-Hochberg-corrected *p*-value.

Food Tested	Compound	Effect direction	Linear model effect size	Raw *p*-value	FDR *p*-value	Compound relevance/reference
Alcohol	Ethyl sulfate	Increase	0.36	1.61E-06	1.91E-03	Established alcohol biomarker[27]
Alcohol	Stachydrine	Increase	0.28	2.40E-05	9.46E-03	Found in some grains, fruit, and other plants[86-88]
Alcohol	Trp-Lys	Decrease	-0.37	1.85E-05	9.46E-03	Dipeptides; unknown relevance
Dessert	3-hydroxy-4-methoxybenzoic acid	Increase	0.33	4.10E-04	4.85E-01	Breakdown product of isovanillin found in many foods[28]
Dessert	Neochlorogenic acid	Increase	0.31	7.66E-03	9.92E-01	Plant polyphenol, abundant in some fruits and berries[29,89]
Fruit	N-methylproline	Increase	0.41	4.31E-08	5.10E-05	Blood serum biomarker of fruit consumption[25]
Fruit	Stachydrine	Increase	0.26	3.24E-05	1.91E-02	Found in some grains, fruit, and other plants[86-88]
Fruit	Ketoisovaleric acid	Increase	0.29	2.34E-04	9.23E-02	Keto acid, ubiquitous metabolic intermediates[90]
Fruit	Betonicine	Increase	0.23	4.65E-04	1.37E-01	Component of citrus fruits[26]
Fruit	4-methyl-2-oxovaleric acid	Increase	0.23	1.95E-03	3.83E-01	Keto acid, ubiquitious metabolic intermediates[90]
Fruit	Hexanoyl-L-carnitine	Decrease	-0.31	3.14E-03	3.83E-01	Acyl-carnitine, prevalent metabolic intermediate[45]

Table 3: Subject Demographics. Detailed exclusion and inclusion criteria are supplied in supplemental methods.

Attribute	Value
Total number of subjects	15
Subjects completing the study	15
Age	Mean 42, range 22-64
Females	8
Males	7
Antibiotic use within past 6 months	2
Underlying medical conditions	0

References

1 Sensoy, I. A review on the food digestion in the digestive tract and the used in vitro models. *Curr Res Food Sci* **4**, 308-319, doi:10.1016/j.crfs.2021.04.004 (2021).

2 Soderstrom, L., Rosenblad, A., Thors Adolfsson, E. & Bergkvist, L. Malnutrition is associated with increased mortality in older adults regardless of the cause of death. *Br J Nutr* **117**, 532-540, doi:10.1017/S0007114517000435 (2017).

3 Kastin, D. A. & Buchman, A. L. Malnutrition and gastrointestinal disease. *Curr Opin Gastroenterol* **18**, 221-228, doi:10.1097/00001574-200203000-00012 (2002).

4 Annunziata, M. L., Caviglia, R., Papparella, L. G. & Cicala, M. Upper gastrointestinal involvement of Crohn's disease: a prospective study on the role of upper endoscopy in the diagnostic work-up. *Dig Dis Sci* **57**, 1618-1623, doi:10.1007/s10620-012-2072-0 (2012).

5 Kovari, B. & Pai, R. K. Upper Gastrointestinal Tract Involvement in Inflammatory Bowel Diseases: Histologic Clues and Pitfalls. *Adv Anat Pathol* **29**, 2-14, doi:10.1097/PAP.0000000000000311 (2022).

6 Tang, Q. *et al.* Current sampling methods for gut microbiota: a call for more precise devices. *Frontiers in cellular and infection microbiology*, 151 (2020).

7 Bloszies, C. S. & Fiehn, O. Using untargeted metabolomics for detecting exposome compounds. *Current Opinion in Toxicology* **8**, 87-92 (2018).

8 Dame, Z. T. *et al.* The human saliva metabolome. *Metabolomics* **11**, 1864-1883 (2015).

9 Hollander, F. The components of the gastric secretion. *American Journal of Digestive Diseases* **3**, 651-654 (1936).

10 Esteller, A. Physiology of bile secretion. *World J Gastroenterol* **14**, 5641-5649, doi:10.3748/wjg.14.5641 (2008).

11 Cortese, N. *et al.* Metabolome of Pancreatic Juice Delineates Distinct Clinical Profiles of Pancreatic Cancer and Reveals a Link between Glucose Metabolism and PD-1(+) Cells. *Cancer Immunol Res* **8**, 493-505, doi:10.1158/2326-6066.CIR-19-0403 (2020).

12 Johansson, M. E., Sjovall, H. & Hansson, G. C. The gastrointestinal mucus system in health and disease. *Nat Rev Gastroenterol Hepatol* **10**, 352-361, doi:10.1038/nrgastro.2013.35 (2013).

13 Hanson, A. D., Henry, C. S., Fiehn, O. & de Crecy-Lagard, V. Metabolite Damage and Metabolite Damage Control in Plants. *Annu Rev Plant Biol* **67**, 131-152, doi:10.1146/annurev-arplant-043015-111648 (2016).

14 Folz, J. S., Shalon, D. & Fiehn, O. Metabolomics analysis of time-series human small intestine lumen samples collected in vivo. *Food Funct* **12**, 9405-9415, doi:10.1039/d1fo01574e (2021).

15 Hoffmann, M. A. *et al.* High-confidence structural annotation of metabolites absent from spectral libraries. *Nat Biotechnol* **40**, 411-421, doi:10.1038/s41587-021-01045-9 (2022).

16 Evans, D. F. *et al.* Measurement of gastrointestinal pH profiles in normal ambulant human subjects. *Gut* **29**, 1035-1041, doi:10.1136/gut.29.8.1035 (1988).

17 Shalon, D. *et al.* Profiling of the human intestinal microbiome and bile acids under physiologic conditions using an ingestible sampling device. *bioRxiv* (2022).

18 Schymanski, E. L. *et al.* Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ Sci Technol* **48**, 2097-2098, doi:10.1021/es5002105 (2014).

19 Djoumbou Feunang, Y. *et al.* ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J Cheminform* **8**, 61, doi:10.1186/s13321-016-0174-y (2016).

20 Rios-Covian, D. *et al.* Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health. *Front Microbiol* **7**, 185, doi:10.3389/fmicb.2016.00185 (2016).

21 Silvester, K. R., Englyst, H. N. & Cummings, J. H. Ileal recovery of starch from whole diets containing resistant starch measured in vitro and fermentation of ileal effluent. *Am J Clin Nutr* **62**, 403-411, doi:10.1093/ajcn/62.2.403 (1995).

22 Ridlon, J. M., Kang, D. J. & Hylemon, P. B. Bile salt biotransformations by human intestinal bacteria. *J Lipid Res* **47**, 241-259, doi:10.1194/jlr.R500013-JLR200 (2006).

23 Lucas, L. N. *et al.* Dominant Bacterial Phyla from the Human Gut Show Widespread Ability To Transform and Conjugate Bile Acids. *mSystems*, e0080521, doi:10.1128/mSystems.00805-21 (2021).

24 Quinn, R. A. *et al.* Global chemical effects of the microbiome include new bile-acid conjugations. *Nature* **579**, 123-129, doi:10.1038/s41586-020-2047-9 (2020).

25 Guertin, K. A. *et al.* Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. *Am J Clin Nutr* **100**, 208-217, doi:10.3945/ajcn.113.078758 (2014).

26 Servillo, L. *et al.* Betaines in fruits of Citrus genus plants. *J Agric Food Chem* **59**, 9410-9416, doi:10.1021/jf2014815 (2011).

27 Hoiseth, G., Morini, L., Polettini, A., Christophersen, A. & Morland, J. Blood kinetics of ethyl glucuronide and ethyl sulphate in heavy drinkers during alcohol detoxification. *Forensic Sci Int* **188**, 52-56, doi:10.1016/j.forsciint.2009.03.017 (2009).

28 Panoutsopoulos, G. I. & Beedham, C. Enzymatic oxidation of vanillin, isovanillin and protocatechuic aldehyde with freshly prepared Guinea pig liver slices. *Cell Physiol Biochem* **15**, 89-98, doi:10.1159/000083641 (2005).

29 Rothwell, J. A. *et al.* Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content. *Database (Oxford)* **2013**, bat070, doi:10.1093/database/bat070 (2013).

30 Ballistreri, G. *et al.* Fruit quality and bioactive compounds relevant to human health of sweet cherry (Prunus avium L.) cultivars grown in Italy. *Food Chem* **140**, 630-638, doi:10.1016/j.foodchem.2012.11.024 (2013).

31 Capitani, D. *et al.* Peach fruit: metabolic comparative analysis of two varieties with different resistances to insect attacks by NMR spectroscopy. *J Agric Food Chem* **61**, 1718-1726, doi:10.1021/jf303248z (2013).

32 Hamoud, A. R., Weaver, L., Stec, D. E. & Hinds, T. D., Jr. Bilirubin in the Liver-Gut Signaling Axis. *Trends Endocrinol Metab* **29**, 140-150, doi:10.1016/j.tem.2018.01.002 (2018).

33 Vitek, L. *et al.* Identification of bilirubin reduction products formed by Clostridium perfringens isolated from human neonatal fecal flora. *J Chromatogr B Analyt Technol Biomed Life Sci* **833**, 149-157, doi:10.1016/j.jchromb.2006.01.032 (2006).

34 Saxerholt, H. *et al.* Influence of antibiotics on the faecal excretion of bile pigments in healthy subjects. *Scand J Gastroenterol* **21**, 991-996, doi:10.3109/00365528608996410 (1986).

35 Balas, L., Feillet-Coudray, C. & Durand, T. Branched Fatty Acyl Esters of Hydroxyl Fatty Acids (FAHFAs), Appealing Beneficial Endogenous Fat against Obesity and Type-2 Diabetes. *Chemistry* **24**, 9463-9476, doi:10.1002/chem.201800853 (2018).

36 Yore, M. M. *et al.* Discovery of a class of endogenous mammalian lipids with anti-diabetic and anti-inflammatory effects. *Cell* **159**, 318-332, doi:10.1016/j.cell.2014.09.035 (2014).

37 Yasuda, S. *et al.* Elucidation of Gut Microbiota-Associated Lipids Using LC-MS/MS and 16S rRNA Sequence Analyses. *iScience* **23**, 101841, doi:10.1016/j.isci.2020.101841 (2020).

38 Xu, Y., Zhu, Y., Li, X. & Sun, B. Dynamic balancing of intestinal short-chain fatty acids: The crucial role of bacterial metabolism. *Trends in Food Science & Technology* **100**, 118-130 (2020).

39 Tsugawa, H. *et al.* A lipidome atlas in MS-DIAL 4. *Nat Biotechnol* **38**, 1159-1163, doi:10.1038/s41587-020-0531-2 (2020).

40 Hou, L. *et al.* Identification and Biosynthesis of Pro-Inflammatory Sulfonolipids from an Opportunistic Pathogen Chryseobacterium gleum. *ACS Chem Biol* **17**, 1197-1206, doi:10.1021/acschembio.2c00141 (2022).

41 Walker, A. *et al.* Sulfonolipids as novel metabolite markers of Alistipes and Odoribacter affected by high-fat diets. *Sci Rep* **7**, 11047, doi:10.1038/s41598-017-10369-z (2017).

42 Kamath, A. V., Darling, I. M. & Morris, M. E. Choline uptake in human intestinal Caco-2 cells is carrier-mediated. *J Nutr* **133**, 2607-2611, doi:10.1093/jn/133.8.2607 (2003).

43 Bhutia, Y. D. & Ganapathy, V. in *Physiology of the gastrointestinal tract* 1063-1086 (Elsevier, 2018).

44 McCloud, E., Ma, T. Y., Grant, K. E., Mathis, R. K. & Said, H. M. Uptake of L-carnitine by a human intestinal epithelial cell line, Caco-2. *Gastroenterology* **111**, 1534-1540, doi:10.1016/s0016-5085(96)70015-x (1996).

45 Reuter, S. E. & Evans, A. M. Carnitine and acylcarnitines: pharmacokinetic, pharmacological and clinical aspects. *Clin Pharmacokinet* **51**, 553-572, doi:10.2165/11633940-000000000-00000

10.1007/BF03261931 (2012).

46 Nilsson, A. & Duan, R. D. Alkaline sphingomyelinases and ceramidases of the gastrointestinal tract. *Chem Phys Lipids* **102**, 97-105, doi:10.1016/s0009-3084(99)00078-x (1999).

47 Zhang, X. *et al.* Widespread protein lysine acetylation in gut microbiome and its alterations in patients with Crohn's disease. *Nat Commun* **11**, 4120, doi:10.1038/s41467-020-17916-9 (2020).

48 Arnaud, A., Ramirez, M., Baxter, J. H. & Angulo, A. J. Absorption of enterally administered N-acetyl-l-glutamine versus glutamine in pigs. *Clin Nutr* **23**, 1303-1312, doi:10.1016/j.clnu.2004.04.004 (2004).

49 Stegink, L. D., Filer, L. J., Jr. & Baker, G. L. Plasma methionine levels in normal adult subjects after oral loading with L-methionine and N-acetyl-L-methionine. *J Nutr* **110**, 42-49, doi:10.1093/jn/110.1.42

(1980).

50 Wanders, R. J., Komen, J. & Kemp, S. Fatty acid omega-oxidation as a rescue pathway for fatty acid oxidation disorders in humans. *FEBS J* **278**, 182-194, doi:10.1111/j.1742-4658.2010.07947.x (2011).

51 Miura, Y. The biological significance of omega-oxidation of fatty acids. *Proc Jpn Acad Ser B Phys Biol Sci* **89**, 370-382, doi:10.2183/pjab.89.370 (2013).

52 Craft, D. L., Madduri, K. M., Eshoo, M. & Wilson, C. R. Identification and characterization of the CYP52 family of Candida tropicalis ATCC 20336, important for the conversion of fatty acids and alkanes to alpha,omega-dicarboxylic acids. *Appl Environ Microbiol* **69**, 5983-5991, doi:10.1128/AEM.69.10.5983-5991.2003 (2003).

53 McKenna, E. J. & Coon, M. J. Enzymatic omega-oxidation. IV. Purification and properties of the omega-hydroxylase of Pseudomonas oleovorans. *J Biol Chem* **245**, 3882-3889 (1970).

54 Behne, M. *et al.* Omega-hydroxyceramides are required for corneocyte lipid envelope (CLE) formation and normal epidermal permeability barrier function. *J Invest Dermatol* **114**, 185-192, doi:10.1046/j.1523-1747.2000.00846.x (2000).

55 Bohn, T. *et al.* Mind the gap-deficits in our knowledge of aspects impacting the bioavailability of phytochemicals and their metabolites–a position paper focusing on carotenoids and polyphenols. *Mol Nutr Food Res* **59**, 1307-1323, doi:10.1002/mnfr.201400745 (2015).

56 Valdes, L. *et al.* The relationship between phenolic compounds from diet and microbiota: impact on human health. *Food Funct* **6**, 2424-2439, doi:10.1039/c5fo00322a (2015).

57 Nagar, E. E., Okun, Z. & Shpigelman, A. Digestive fate of polyphenols: Updated view of the influence of chemical structure and the presence of cell wall material. *Current Opinion in Food Science* **31**, 38-46 (2020).

58 Day, A. J. *et al.* Dietary flavonoid and isoflavone glycosides are hydrolysed by the lactase site of lactase phlorizin hydrolase. *FEBS Lett* **468**, 166-170, doi:10.1016/s0014-5793(00)01211-4 (2000).

59 Parada, J. & Aguilera, J. M. Food microstructure affects the bioavailability of several nutrients. *J Food Sci* **72**, R21-32, doi:10.1111/j.1750-3841.2007.00274.x (2007).

60 Saura-Calixto, F. *et al.* Proanthocyanidin metabolites associated with dietary fibre from in vitro colonic fermentation and proanthocyanidin metabolites in human plasma. *Mol Nutr Food Res* **54**, 939-946, doi:10.1002/mnfr.200900276 (2010).

61 Tydeman, E. A. *et al.* Effect of carrot (Daucus carota) microstructure on carotene bioaccessibility in the upper gastrointestinal tract. 2. In vivo digestions. *J Agric Food Chem* **58**, 9855-9860, doi:10.1021/jf1010353 (2010).

62 Kuijsten, A., Arts, I. C., van't Veer, P. & Hollman, P. C. The relative bioavailability of enterolignans in humans is enhanced by milling and crushing of flaxseed. *J Nutr* **135**, 2812-2816, doi:10.1093/jn/135.12.2812 (2005).

63 Jenab, M. & Thompson, L. U. The influence of flaxseed and lignans on colon carcinogenesis and beta-glucuronidase activity. *Carcinogenesis* **17**, 1343-1348, doi:10.1093/carcin/17.6.1343 (1996).

64 Hossain, G. S. *et al.* L-Amino acid oxidases from microbial sources: types, properties, functions, and applications. *Appl Microbiol Biotechnol* **98**, 1507-1515, doi:10.1007/s00253-013-5444-2 (2014).

65 Deng, X., Wood, P. G., Eagon, P. K. & Whitcomb, D. C. Chronic alcohol-induced alterations in the pancreatic secretory control mechanisms. *Dig Dis Sci* **49**, 805-819, doi:10.1023/b:ddas.0000030093.25897.61 (2004).

66 Judak, L. *et al.* Ethanol and its non-oxidative metabolites profoundly inhibit CFTR function in pancreatic epithelial cells which is prevented by ATP supplementation. *Pflugers Arch* **466**, 549-562, doi:10.1007/s00424-013-1333-x (2014).

67 Kotorman, M., Laczko, I., Szabo, A. & Simon, L. M. Effects of Ca2+ on catalytic activity and conformation of trypsin and alpha-chymotrypsin in aqueous ethanol. *Biochem Biophys Res Commun* **304**, 18-21, doi:10.1016/s0006-291x(03)00534-5 (2003).

68 Tassaneeyakul, W. *et al.* Caffeine metabolism by human hepatic cytochromes P450: contributions of 1A2, 2E1 and 3A isoforms. *Biochem Pharmacol* **47**, 1767-1776, doi:10.1016/0006-2952(94)90304-2 (1994).

69 Gummadi, S. N., Bhavya, B. & Ashok, N. Physiology, biochemistry and possible applications of microbial caffeine degradation. *Appl Microbiol Biotechnol* **93**, 545-554, doi:10.1007/s00253-011-3737-x (2012).

70 Summers, R. M., Mohanty, S. K., Gopishetty, S. & Subramanian, M. Genetic characterization of caffeine degradation by bacteria and its potential applications. *Microb Biotechnol* **8**, 369-378, doi:10.1111/1751-7915.12262 (2015).

71 Blanchard, J. & Sawers, S. J. The absolute bioavailability of caffeine in man. *Eur J Clin Pharmacol* **24**, 93-98, doi:10.1007/BF00613933 (1983).

72 Arnaud, M. J. Components of coffee. *Caffeine, coffee, and health* **43** (1993).

73 Beach, C. A., Mays, D. C., Sterman, B. M. & Gerber, N. Metabolism, distribution, seminal excretion and pharmacokinetics of caffeine in the rabbit. *J Pharmacol Exp Ther* **233**, 18-23 (1985).

74 Vitek, L. & Ostrow, J. D. Bilirubin chemistry and metabolism; harmful and protective aspects. *Curr Pharm Des* **15**, 2869-2883, doi:10.2174/138161209789058237 (2009).

75 Liberati-Cizmek, A. M. *et al.* Analysis of Fatty Acid Esters of Hydroxyl Fatty Acid in Selected Plant Food. *Plant Foods Hum Nutr* **74**, 235-240, doi:10.1007/s11130-019-00728-8 (2019).

76 Gibson, G., Cummings, J. & Macfarlane, G. Growth and activities of sulphate-reducing bacteria in gut contents of healthy subjects and patients with ulcerative colitis. *FEMS microbiology letters* **86**, 103-111 (1991).

77 Maeda, J. *et al.* Inhibitory effects of sulfobacin B on DNA polymerase and inflammation. *Int J Mol Med* **26**, 751-758, doi:10.3892/ijmm_00000522 (2010).

78 Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A. & Schwudke, D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J Lipid Res* **49**, 1137-1146, doi:10.1194/jlr.D700041-JLR200 (2008).

79 Nusbaum, D. J. *et al.* Gut microbial and metabolomic profiles after fecal microbiota transplantation in pediatric ulcerative colitis patients. *FEMS Microbiol Ecol* **94**, doi:10.1093/femsec/fiy133 (2018).

80 Barupal, D. K. *et al.* A Comprehensive Plasma Metabolomics Dataset for a Cohort of Mouse Knockouts within the International Mouse Phenotyping Consortium. *Metabolites* **9**, doi:10.3390/metabo9050101 (2019).

81 Celis, A. I. *et al.* Optimization of the 16S rRNA sequencing analysis pipeline for studying in vitro communities of gut commensals. *iScience* **25**, 103907, doi:10.1016/j.isci.2022.103907 (2022).

82 Team, R. C. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *http://www. R-project. org/* (2013).

83 Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581-583, doi:10.1038/nmeth.3869 (2016).

84 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590-596, doi:10.1093/nar/gks1219 (2013).

85 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. B* **57**, 289-300 (1995).

86 Cheng, F. *et al.* A review of pharmacological and pharmacokinetic properties of stachydrine. *Pharmacol Res* **155**, 104755, doi:10.1016/j.phrs.2020.104755 (2020).

87 Heinzmann, S. S. *et al.* Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. *Am J Clin Nutr* **92**, 436-443, doi:10.3945/ajcn.2010.29672 (2010).

88 Weston, R. J. Analysis of cereals, malted foods and dried legumes for N-nitrosodimethylamine. *Journal of the Science of Food and Agriculture* **35**, 782-786 (1984).

89 Möller, B. & Herrmann, K. Quinic acid esters of hydroxycinnamic acids in stone and pome fruit. *Phytochemistry* **22**, 477-481 (1983).

90 Luo, Z., Yu, S., Zeng, W. & Zhou, J. Comparative analysis of the chemical and biochemical synthesis of keto acids. *Biotechnol Adv* **47**, 107706, doi:10.1016/j.biotechadv.2021.107706 (2021).

Figures

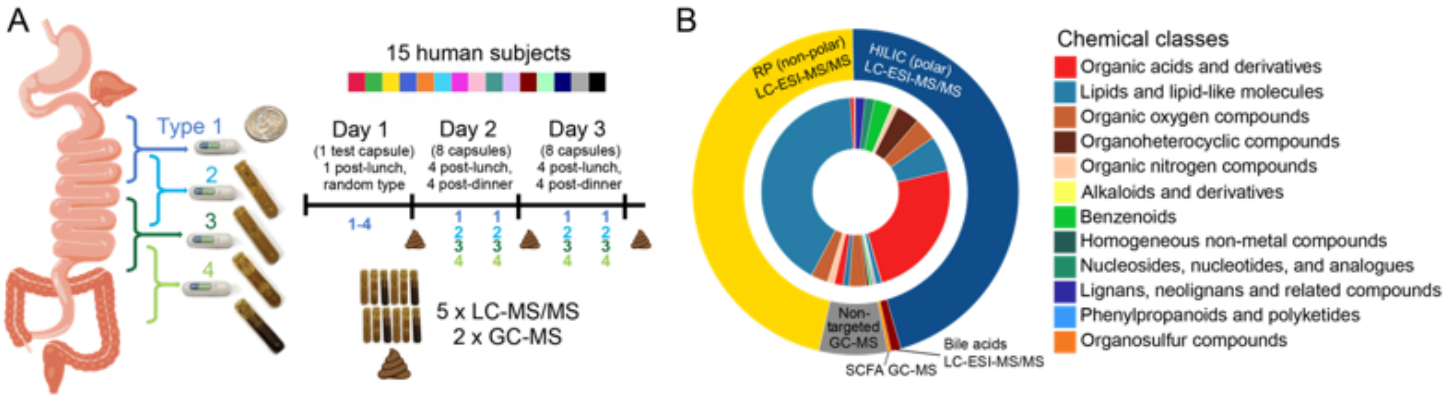

Figure 1

Study design and metabolomics summary. (A) Study design for upper intestinal tract investigation. Four designs of an intestinal sampling device were used to sample the proximal to distal upper intestinal region. Fifteen human volunteers swallowed at least 16 capsules over two days after lunch and after dinner after an initial test on day 1. Capsules were retrieved and analyzed by targeted and non-targeted LC-MS/MS and GC-MS methods. (B) Identified metabolites from the different analytical methods used to analyze samples including non-targeted reverse phase (RP) liquid chromatography (LC) with electrospray ionization (ESI) and tandem high resolution mass spectrometry (MS/MS), hydrophilic interaction liquid chromatography (HILIC) ESI-MS/MS, non-targeted gas chromatography (GC) mass spectrometry (MS), targeted quantification of short chain fatty acids (SCFA) by GC-MS, and targeted bile acid quantification by LC-ESI-MS/MS. Chemical class fractions are included based on automated ClassyFire chemical classification.

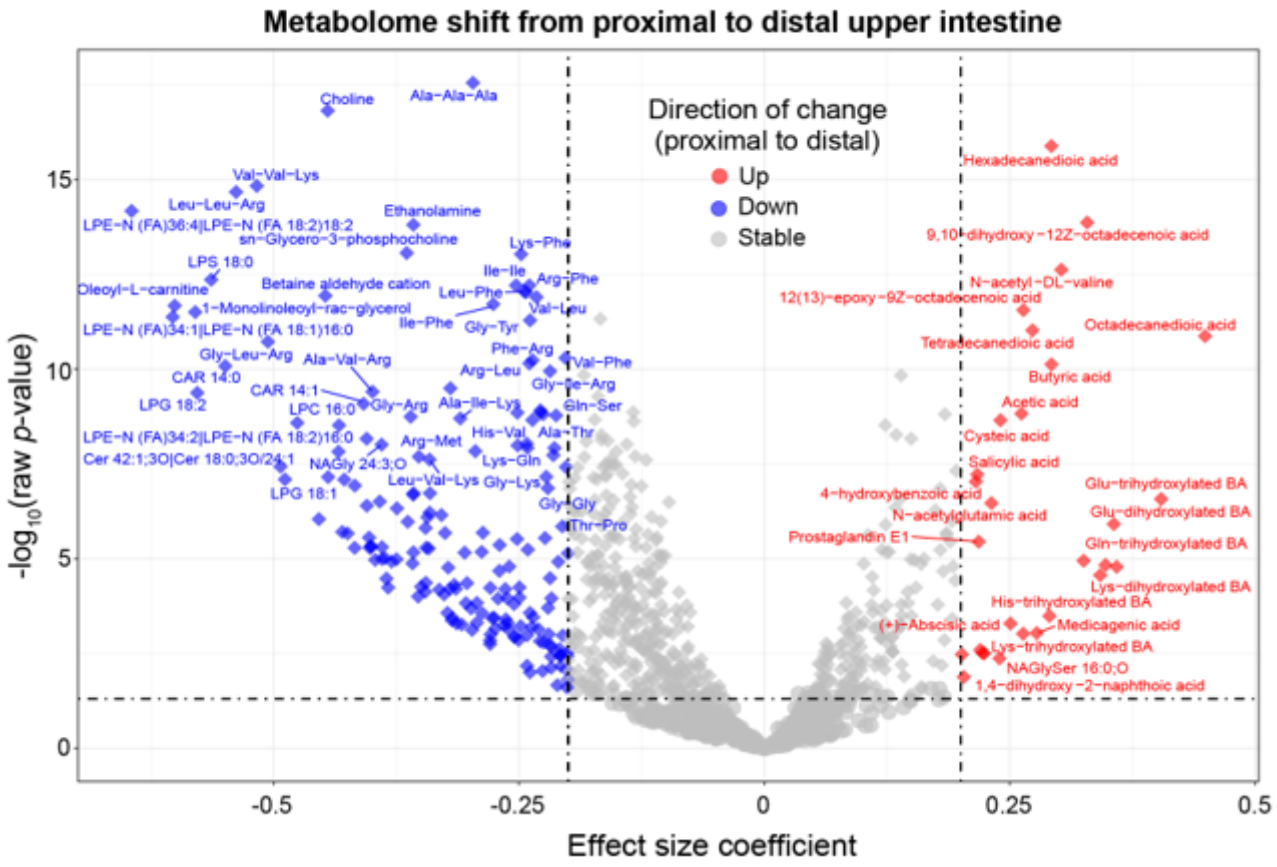

Figure 2

Comparison of proximal and distal upper intestinal metabolite levels. Significance was calculated by linear mixed effect model (LMM). Significance of $p < 0.05$ is delimited by the lower dashed horizontal line. Circle shape for a metabolite indicates non-significance after FDR correction and diamond shape indicates significance ($p<0.05$) after false discovery rate (FDR) correction (n=1,182). Metabolites detected in >50% of intestinal samples were included in this analysis (n=1,182). Effect size coefficient is the slope estimated by LMM with positive (negative) coefficient meaning higher (lower) in the distal compared to proximal upper intestine. Vertical dashed lines are ±0.2 effect size coefficient.

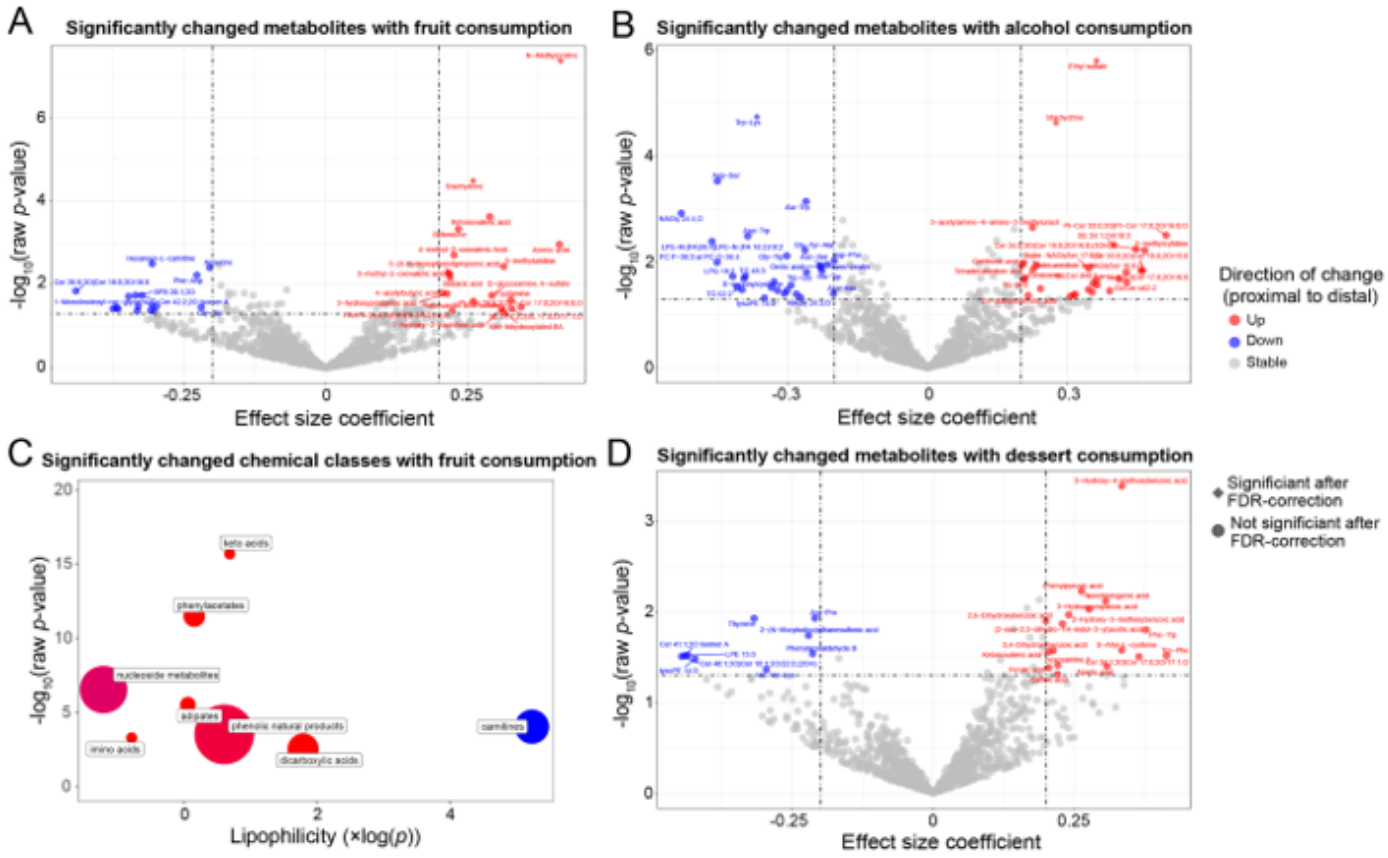

Figure 3

Intestinal metabolite association with food types. Volcano plots show significance of each metabolite to food type (A) fruit, (B) alcohol, and (D) dessert calculated by linear mixed effect model (LMM). Consumption is considered food eaten within 6 hours of swallowing sample devices. Significance of $p < 0.05$ is delimited by the lower dashed horizontal line. Circle shape for a metabolite indicates non-significance after false discovery rate (FDR) correction and diamond shape indicates significance (p<0.05) after FDR correction (n=1,182). Metabolites detected in >50% of intestinal samples were included in this analysis. Effect size coefficient is the slope estimate calculated by LMM, with positive (negative) coefficient meaning the metabolite was higher (lower) after food consumption. Vertical dashed lines are ±0.2 effect size coefficient. (C) Chemical enrichment statistics (ChemRICH) analysis revealed significant chemical classes after fruit consumption visualized by separating classes by chemical lipophilicity (logP) and chemical class significance level of -$\log_{10}$(p-value). Red circle indicates the chemical class increased after fruit consumption and blue indicates the chemical class decreased after fruit consumption. The size of circles indicates the size of the chemical class.

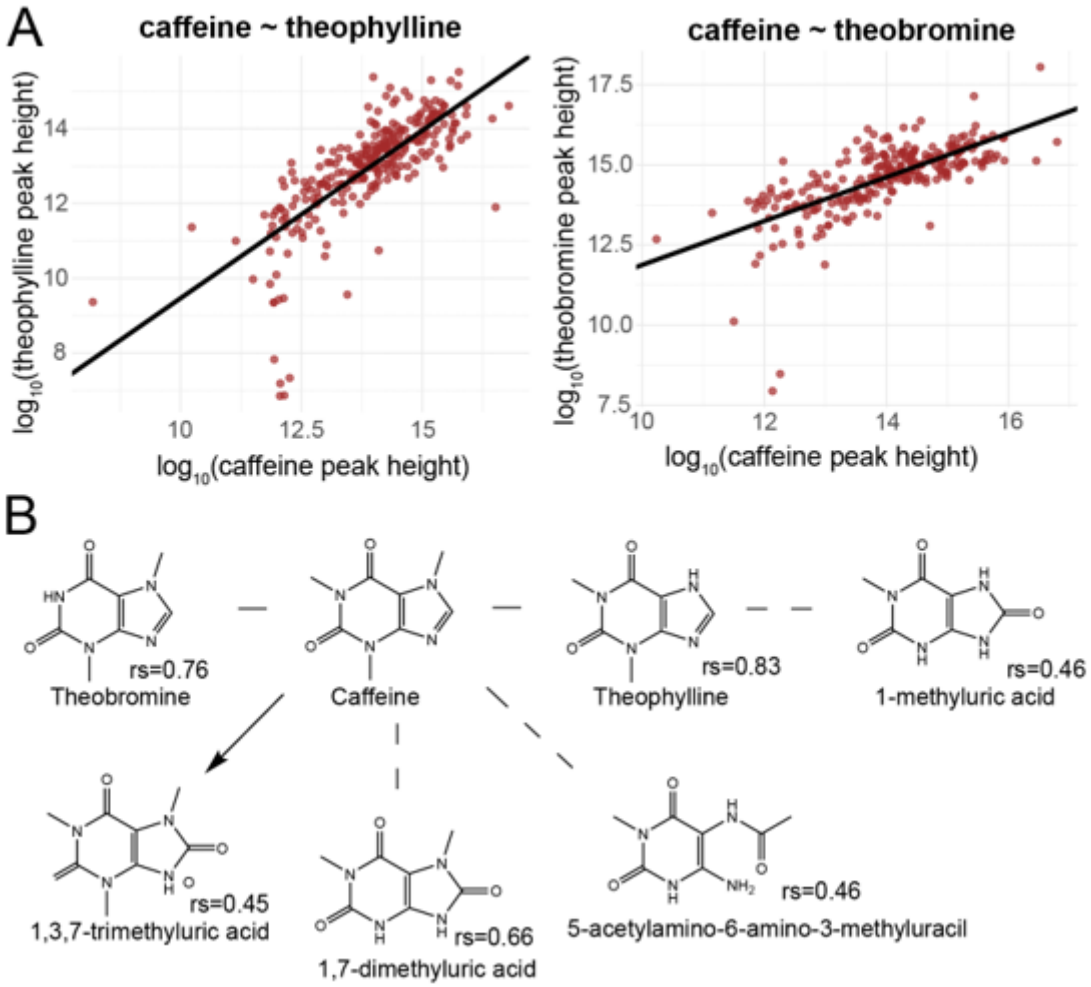

Figure 4

Caffeine and related metabolites. (A) Plots of $\log_{10}$(peak height) of caffeine to theophylline, and caffeine to theobromine for each sample where both metabolites were detected. (B) Chemical diagram of caffeine and known metabolic pathways with detected metabolite structures shown and Spearman rank correlation coefficient (rs) reported for each structure. All Spearman rank correlations were significant with $p < 1.0\times10^{-13}$.

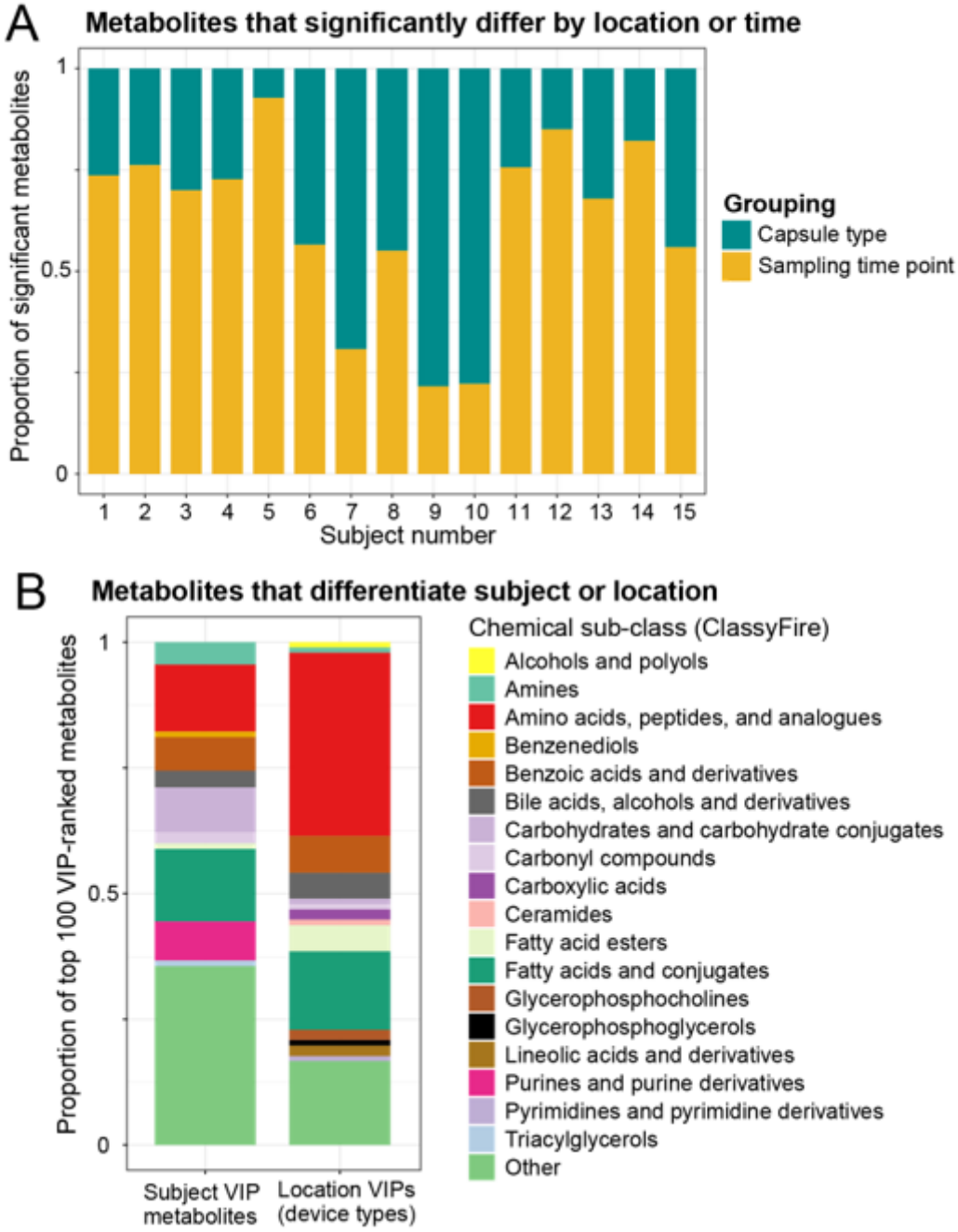

Figure 5

Temporal and personal variation compared to spatial variation in the intestinal metabolome. (A) The number of metabolites that were significantly different between intestinal regions (device types) or between meals (sampling time points of four CapScan devices) were calculated for each subject by analysis of variance (ANOVA). Only metabolites detected in >50% of samples for each subject were used for this analysis. Non-FDR-corrected $p < 0.05$ was used as significance value cutoff. (B) Multivariate discriminant analysis (PLS-DA) was performed to find the metabolites that were most important to distinguish between different subjects, or to distinguish between different regions (device types). The top 100 metabolites most important for distinguishing these groups are ranked by variable importance in

project score (VIP) and are categorized by chemical subclass. Chemical subclasses with <3 metabolites are reported as “Other”.

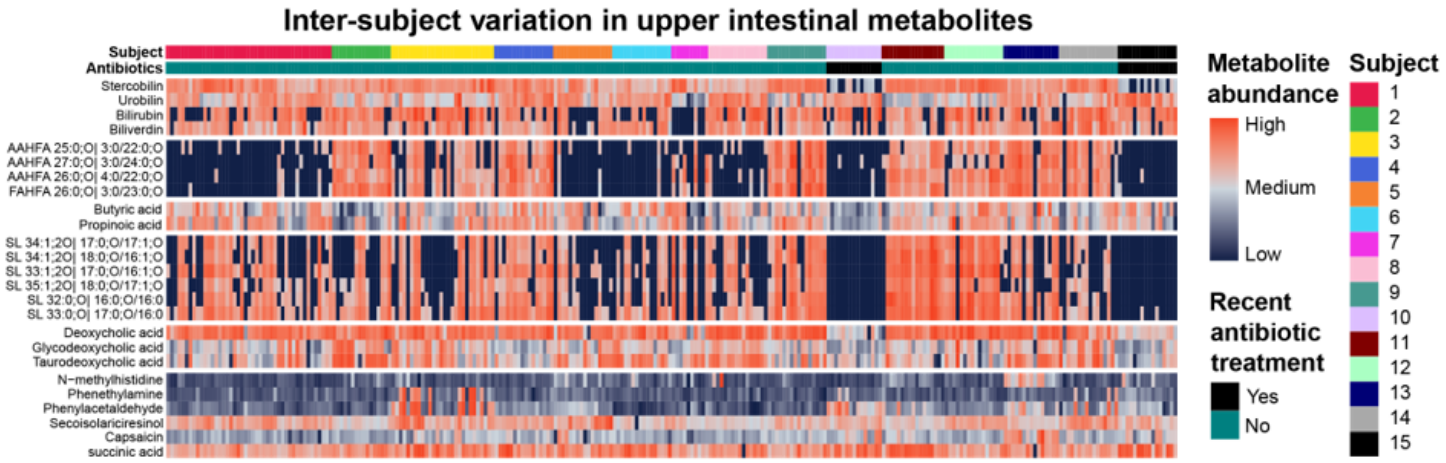

Figure 6

Heatmap of metabolites with strong inter-subject differences. Metabolites include bile pigments, fatty acid esters of hydroxy fatty acids (FAHFAs and AAHFAs), short chain fatty acids, sulfonolipids (SLs), and secondary bile acids. Samples are organized by subject and antibiotic consumption is indicated for the two subjects that consumed antibiotics 1 and 5 months prior to this study. Color of heatmap ranges from low (blue) to high (red) of metabolite abundance (peak height) or concentration (ng/mL) for bile acids. Minimum and maximum values are used to set color scale for each metabolite (each row).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FolzetalNatMetabSupplementaryTables.xlsx
- FolzetalNatMetabSUPPLEMENTALINFO09122022.docx