

## UC Davis

### UC Davis Previously Published Works

**Title**

Human metabolome variation along the upper intestinal tract

**Permalink**

<https://escholarship.org/uc/item/8kq2w6q8>

**Journal**

Nature Metabolism, 5(5)

**ISSN**

2522-5812

**Authors**

Folz, Jacob  
Culver, Rebecca Neal  
Morales, Juan Montes  
et al.

**Publication Date**

2023-05-01

**DOI**

10.1038/s42255-023-00777-z

Peer reviewed

# Human metabolome variation along the upper intestinal tract

Received: 24 September 2022

Accepted: 3 March 2023

Published online: 10 May 2023

 Check for updates

Jacob Folz<sup>1</sup>, Rebecca Neal Culver<sup>2</sup>, Juan Montes Morales<sup>1</sup>, Jessica Grembi<sup>3</sup>, George Triadafilopoulos<sup>4</sup>, David A. Relman<sup>3,5,6,7</sup>, Kerwyn Casey Huang<sup>5,6,8</sup>, Dari Shalon<sup>9</sup> & Oliver Fiehn<sup>1</sup>✉

Most processing of the human diet occurs in the small intestine. Metabolites in the small intestine originate from host secretions, plus the ingested exposome<sup>1</sup> and microbial transformations. Here we probe the spatiotemporal variation of upper intestinal luminal contents during routine daily digestion in 15 healthy male and female participants. For this, we use a non-invasive, ingestible sampling device to collect and analyse 274 intestinal samples and 60 corresponding stool homogenates by combining five mass spectrometry assays<sup>2,3</sup> and 16S rRNA sequencing. We identify 1,909 metabolites, including sulfonolipids and fatty acid esters of hydroxy fatty acids (FAHFA) lipids. We observe that stool and intestinal metabolomes differ dramatically. Food metabolites display 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 account for the largest inter-individual differences. Notably, two individuals who had taken antibiotics within 6 months before sampling show large variation in levels of bioactive FAHFAs and sulfonolipids and other microbially related metabolites. From inter-individual variation, we identify *Blautia* species as a candidate to be involved in FAHFA metabolism. In conclusion, non-invasive, in vivo sampling of the human small intestine and ascending colon under physiological conditions reveals links between diet, host and microbial metabolism.

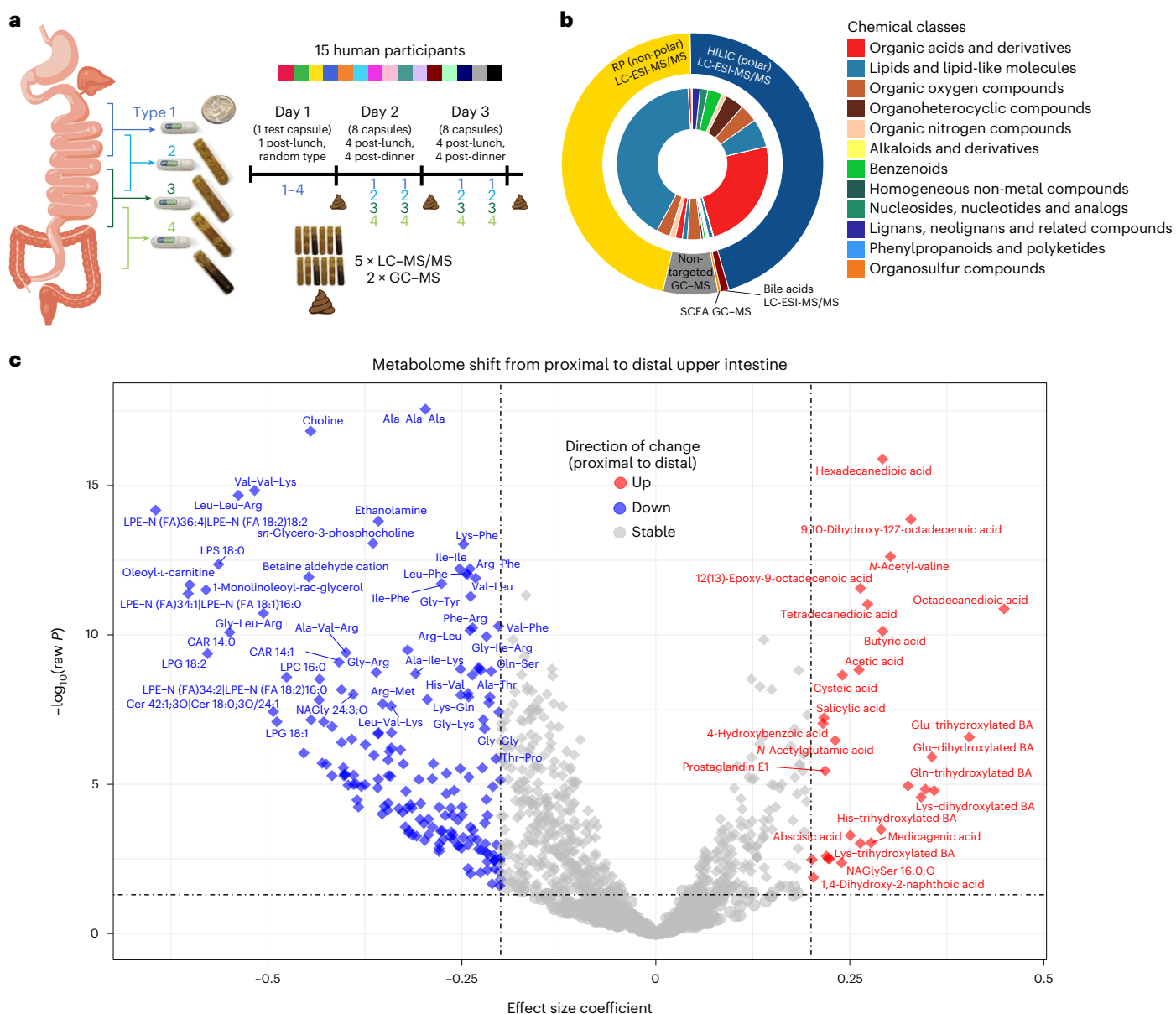
We aimed to comprehensively study metabolomic differences among luminal samples from the upper intestinal tract of 15 healthy individuals to better understand the extent of spatial and temporal variation and to gauge the prospects of integrating metabolome and microbiome data. In a related companion publication<sup>4</sup>, we use these devices to study variation along the gut in microbiota composition, prophage induction, the host proteome and microbial modification of bile acids.

Volunteers swallowed sets of four sampling devices per sampling time point. These ingestible sampling devices consisted of a collapsed collection bladder capped by a one-way valve in a capsule with a pH-sensitive coating. The four types of devices differed only in their enteric coating, which dissolved at pH 5.5 (type 1), pH 6 (type 2) and pH 7.5 (types 3 and 4) (Fig. 1a). The thickness and pH responsiveness of the coating enabled sampling at specific locations of the intestinal tract after gastric

<sup>1</sup>West Coast Metabolomics Center, University of California, Davis, CA, USA. <sup>2</sup>Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. <sup>3</sup>Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. <sup>4</sup>Silicon Valley Neurogastroenterology and Motility Center, Mountain View, CA, USA. <sup>5</sup>Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA.

<sup>6</sup>Chan Zuckerberg Biohub, San Francisco, CA, USA. <sup>7</sup>Infectious Diseases Section, Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, USA.

<sup>8</sup>Department of Bioengineering, Stanford University, Stanford, CA, USA. <sup>9</sup>Envivo Bio, San Francisco, CA, USA. ✉e-mail: [ofiehn@ucdavis.edu](mailto:ofiehn@ucdavis.edu)



**Fig. 1 | Comparison of proximal and distal upper intestinal metabolite levels reveals significant differences in a wide range of compounds.**

**a**, Study design for upper intestinal tract investigation. Four types of an intestinal sampling device were used to sample the proximal to distal upper intestines. Fifteen human participants swallowed at least 16 devices over 2 d after lunch and after dinner after an initial test on day 1. Devices were retrieved and analysed by targeted and non-targeted LC-MS/MS and GC-MS methods. **b**, Identified metabolites from the five metabolome assays used to analyse samples. Chemical class fractions are included based on automated ClassyFire chemical

classification. **c**, Significance of differences between upper intestinal tract regions was calculated using LMM. Horizontal dashed-dotted line represents the significance threshold  $P < 0.05$  ( $n = 1,182$  metabolites). Circles indicate non-significance and diamond shapes indicate significance ( $P < 0.05$ ) after FDR correction. Only 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 indicating higher (lower) levels in the distal compared to proximal upper intestine. Vertical dashed-dotted lines are  $\pm 0.2$  effect size coefficient.

emptying. 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  $\mu\text{l}$  of luminal contents through vacuum suction. The one-way valve prevented loss of sample and contamination from downstream fluids. Stool samples were frozen at  $-20^\circ\text{C}$  and all devices were recovered from the stool before analysis. Liquid contents were retrieved from devices using hypodermic needles. Aliquots of the raw sample were used for 16S ribosomal RNA microbiome analyses and the supernatants from centrifuged samples were used for metabolomic studies. Here, we perform a meticulous analysis of the metabolome

in the same samples, reporting metabolites never before detected in human samples, key biomarkers of diet and comparison of chemical profiles across and within participants (Supplementary Tables 1 and 2).

The measured pH of the luminal contents for device types 1 through 4 was consistent with the expected pH gradient across the intestinal tract<sup>4,5</sup>, covering the duodenum, jejunum, ileum and ascending colon (Fig. 1a). The pH in type 1 and 2 devices was significantly different from type 3 and 4 devices (Extended Data Fig. 1a and Supplementary Table 3; Wilcoxon two-way rank-sum test,  $P = 2.4 \times 10^{-14}$ ), whereas pH was not significantly different between type 1 and type 2 devices or between type 3 and type 4 devices (Extended Data Fig. 1a).

We therefore associated type 1/2 and 3/4 devices with proximal (duodenum and jejunum) and distal (ileum and ascending colon) regions of the upper intestinal tract, respectively.

We used five mass spectrometry assays to analyse the luminal contents captured by 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 Metabolomics Standards Initiative confidence levels 1–3 (Supplementary Table 1)<sup>6</sup>, including 155 internal standards used for quality control (QC) and quantification purposes. Additionally, >12,000 unknown chromatographic features were reliably detected above the level of method blanks (Supplementary Table 2). Using ClassyFire software<sup>7</sup>, structurally annotated metabolites fell into 61 chemical subclasses (Supplementary Table 1). Two untargeted high-resolution liquid chromatography (LC) MS/MS assays focusing on hydrophilic and lipophilic metabolites yielded most of the annotated compounds, with 1,612 identifications. Untargeted gas chromatography (GC)–MS added 119 primary metabolites, supplemented by targeting six short-chain fatty acids (SCFAs) and a targeted LC–MS/MS assay for 17 bile acids (Fig. 1b). QC analysis of total metabolic variance revealed separation of stool and intestinal samples, with strong clustering of pooled quality control samples (Extended Data Fig. 1b).

Metabolome results revealed notable differences between stool and intestinal samples (Extended Data Fig. 2) and among the intestinal tract samples (Extended Data Fig. 3). To uncover spatial differences across the intestine, we applied linear mixed-effect models (LMMs) that accounted for sampling location (proximal or distal) as well as other variables (Supplementary Tables 3 and 4). 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 (false discovery rate (FDR)  $P < 0.05$ ; LMM) (Fig. 1c and Supplementary Table 4), 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 (Fig. 1c). Known microbially generated chemicals including SCFAs<sup>8,9</sup>, secondary bile acids<sup>10</sup> and some microbially conjugated bile acids<sup>11,12</sup>, increased from the proximal to distal upper intestine (Extended Data Table 1 and Fig. 1c). Of the 11 detected acetylated amino acids, 7 increased from the proximal to distal upper intestine (raw  $P < 0.05$ ; LMM) (Extended Data Table 1 and Fig. 1c). 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 (Supplementary File 1). 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) (Extended Data Fig. 4).

To interrogate general metabolic differences between locations, we used chemical enrichment statistics. Di- and tripeptides were among the most significantly decreased classes from the proximal to distal upper intestine (Extended Data Table 1 and Extended Data Fig. 5). Of the 333 di- and tripeptides measured, 262 significantly decreased in abundance from the proximal to distal upper intestine (raw  $P < 0.05$ ; LMM) (Extended Data Table 1). Sugars, sugar alcohols, nucleosides, carnitines and ceramides also exhibited significantly higher levels in proximal intestinal tract samples compared to distal samples (Extended Data Table 1 and Extended Data Fig. 5). These spatial differences in the intestine reflect classic digestion and absorption<sup>13</sup> of di- and tripeptides<sup>14</sup> and acylcarnitines<sup>15,16</sup>, as well as ceramides that are hydrolyzed to sphingosine and free fatty acids before intestinal uptake<sup>17</sup>. In contrast, SCFAs exhibited increased levels in distal regions (Extended Data Table 1 and Fig. 1c), likely due to their production by microbes<sup>8,9</sup>. Acetylated

amino acids, which have been associated with Crohn's disease<sup>18</sup>, were also at higher levels in the distal compared to proximal upper intestine (Extended Data Table 1 and Fig. 1c), possibly due to slower absorption of acetylated compared to non-acetylated amino acids<sup>19,20</sup>. Bile acids are transformed extensively by microbes and levels of secondary bile acids increased along the intestine<sup>4</sup>. These observations support the notion that the capsule devices sampled from the intended locations. Although average pH levels across devices of a given type also followed the expected trends across the upper intestinal tract, the observed within-person variation in pH for each device type did not correlate with metabolomic changes across distal versus proximal regions.

We measured 28 phenolic metabolites that increased from the proximal to distal upper intestine (Supplementary Table 4). These trends are likely caused by a combination of factors, including enzymatic transformation<sup>21,22</sup> such as deglycosylation and delayed breakdown of plant cells and cell-wall components by microbial enzymes<sup>23–25</sup>. For example, the flaxseed-affiliated lignan secoisolariciresinol was most significantly enriched in distal compared to proximal samples, likely due to both bioavailability<sup>26</sup> and deglycosylation<sup>27</sup>.

Dicarboxylic acids also increased in concentration from the proximal to distal upper intestine (Extended Data Table 1 and Fig. 1c). In fact, three of the top six most significantly increased metabolites between the proximal and distal upper intestine were dicarboxylic acids (hexadecanedioic acid, tetradecanedioic acid and octadecanedioic acid) (Fig. 1c). Dicarboxylic acids are generated during catabolism (omega oxidation) of fatty acids, which occurs in human cells<sup>28</sup>, plants<sup>29</sup> and microbes<sup>30,31</sup>. The lead compound, hexadecanedioic acid, was most strongly correlated with other dicarboxylic acids, plant metabolites, bile acids and known microbially produced compounds (Supplementary Table 5). Epithelial cells contain omega-hydroxylated lipids essential to maintain epithelium barrier function<sup>32</sup> 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.

The chemical profiles of intestinal samples differed substantially from those of stool (Extended Data Fig. 2). Thirty-one metabolites were >100 times more abundant on average in the intestine compared to stool. These metabolites consisted of glycinated lipids, sugars, plant natural products, carnitines, microbially conjugated bile acids and *S*-succinylcysteine (Supplementary Table 6). Peptides were also generally at much lower levels in stool samples compared to intestinal samples, especially when compared to the proximal intestine (Extended Data Fig. 2). We also identified >100 metabolites that were >100 times more abundant in stool compared to intestinal samples (Supplementary Table 6); these metabolites were mostly polar lipids such as phosphatidylethanolamines, phosphatidylinositols and phosphatidylglycerols, as well as specific FAHFs. The high abundance of membrane lipids in stool samples is likely due to the high amount of bacterial cell material in stool compared to luminal samples from the upper intestine.

Next, we used LMM to test for associations of food intake logs recorded by the participants 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 h before swallowing capsule devices (Supplementary Table 3). 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 1, Fig. 2a–d and Supplementary Table 4). Despite the small size of this study with 15 participants, we were able to validate a range of dietary biomarkers that were previously found in blood and correlated with fruit<sup>33</sup> and alcohol<sup>34</sup> consumption, as well as discover other biomarkers not previously identified.

Using effect size differences of  $\pm 0.2$  and raw  $P < 0.05$ , fruit consumption was significantly associated with 20 compounds at increased concentration and 17 metabolites at decreased concentration (Fig. 2a).

**Table 1 | Participant characteristics**

Attribute	Value
Total number of participants	15
Participants 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
Body mass index	Mean 23, range 19–31

Detailed exclusion and inclusion criteria are supplied in Supplementary Table 3.

Some metabolites were directly linked to fruit consumption even with strict FDR-corrected  $P < 0.05$  (Fig. 2a), including *N*-methylproline and stachydrine, both of which were previously reported as fruit consumption biomarkers for blood plasma in non-controlled dietary studies<sup>33</sup>. Betonicine, a known component of fruit juice<sup>35</sup>, also increased after fruit consumption at raw  $P < 0.05$  (Fig. 2a) 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  $P < 0.05$  (Fig. 2a). Metabolites are not independent of one another, but rather are linked via food compositions and microbial and enzymatic pathways. Therefore, we used ChemRICH chemical set enrichment statistics to identify significantly altered clusters of metabolites (Fig. 2c). This strategy revealed keto acids as the chemical class with the most significant response to fruit (Fig. 2a,c), highlighting keto acids as a fruit biomarker in the human gut. Keto acids are formed from enzymatic deamination of amino acids, carried out in part by gut bacteria<sup>36</sup>. Notably, ChemRICH also revealed that typical fruit ingredients like phenylacetates and phenolic natural products were positively associated with fruit intake (Fig. 2c).

Alcohol consumption was most significantly associated with ethyl sulfate (FDR  $P < 0.05$ ), a known plasma biomarker of alcohol consumption (Fig. 2b)<sup>34</sup>. Stachydrine was linked with both fruit and alcohol consumption (FDR  $P < 0.05$ ). Trp-Lys significantly decreased with alcohol consumption after FDR correction (Fig. 2b). In total, 40 di- and tripeptides decreased with alcohol consumption (raw  $P < 0.05$ ) with a ChemRICH cluster  $P = 8.8 \times 10^{-18}$  (Supplementary Table 4 and Supplementary Figs. 1 and 2). The decrease in di- and tripeptides after alcohol consumption suggested a decrease in total protease activity, possibly due to impaired pancreatic secretion<sup>37,38</sup> rather than direct inhibition because trypsin and chymotrypsin are active even in 20% ethanol solution<sup>39</sup>. ‘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  $P < 0.05$ ) (Fig. 2d). These compounds are metabolic intermediates in the breakdown of vanillin and isovanillin<sup>40</sup>. Neochlorogenic acid was also significantly associated with dessert (raw  $P < 0.05$ ). Neochlorogenic acid is present in a variety of fruits and berries<sup>41</sup>, including cherries<sup>42</sup> and peaches<sup>43</sup>. Other food types that were included in the mixed-effect model also had significantly associated metabolites (Supplementary Table 4 and Supplementary Figs. 1 and 2).

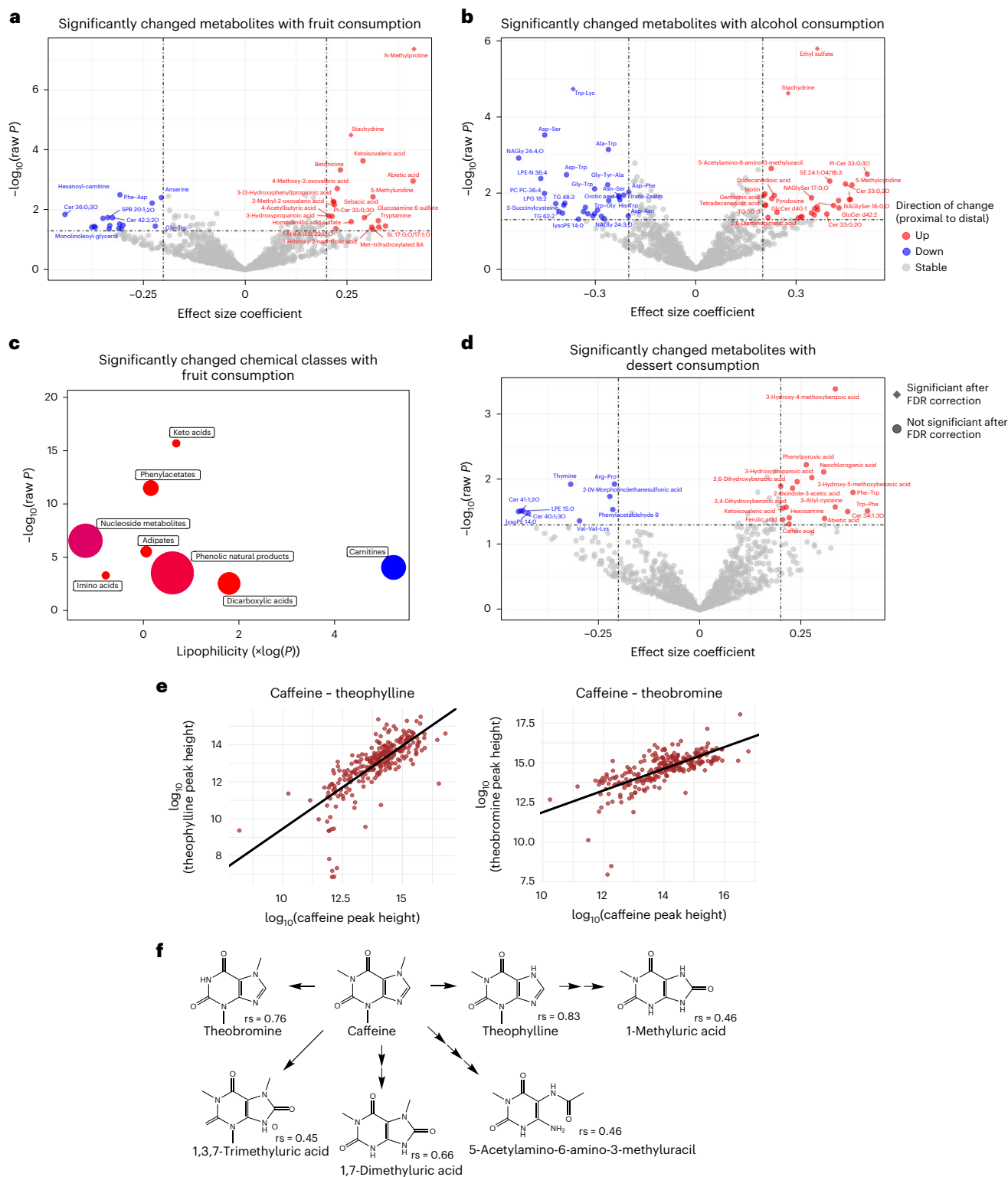
Caffeine was detected in the majority of samples (Supplementary Table 1). Of note, caffeine was not significantly associated with coffee or tea consumption during the experimental timeframe (FDR  $P = 0.87$ ) (Supplementary Table 4), most likely because caffeine is absorbed rapidly within 1 h of oral intake and has a mean half-life of 4.5 h (range 2.7–9.9 h) in the bloodstream<sup>44</sup>; 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<sup>45–47</sup>, including theophylline

and theobromine (Fig. 2e,f). Caffeine is metabolized and excreted through several routes, including urine<sup>48</sup> and bile<sup>49</sup>. Bile is the expected origin of caffeine measured in this study, as multiple hours passed between beverage consumption and device-sampling events. While theobromine is known to be present in chocolate, it did not associate with dessert consumption, only with caffeine metabolism. Hence, upper intestinal tract metabolite correlations may enable reconstruction of microbial and enzymatic pathways of exposome metabolism. Dedicated studies across a diverse population using dietary interventions would be needed to associate specific food biomarkers. In addition, as the devices largely preserve bacterial viability<sup>4</sup>, cultured isolates obtained from the devices could be used to confirm individual food metabolome-bacterial interactions.

Dietary metabolites were associated with temporal differences during the 2 d and four sampling time points of this study. To investigate whether sampling time or sampling region had a larger impact on upper intestinal metabolites, we used analysis of variance (ANOVA) to calculate the number of metabolites that significantly differed between the four device types by participant, or between the four sampling time points (after each meal) by participant. 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 participants had more statistically different metabolites between meals (time points) than between intestinal regions (device types) (Extended Data Fig. 6a,b). A closer inspection of the compound classes that contributed to these differences found that di- and tripeptides (within the chemical class of carboxylic acids) were the largest chemical class that distinguished between device types, representing >70% of all significantly different metabolites in five participants and >40% for another seven participants (Extended Data Fig. 7a). For metabolites that differentiated sampling time points, sugars (organooxygen compounds) were enriched in 13 of 15 participants (Extended Data Fig. 7b). Similarly, more significantly different imidazopyrimidines, indoles and isoflavonoids were found to distinguish sampling time points than intestinal regions (Extended Data Fig. 7). 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.

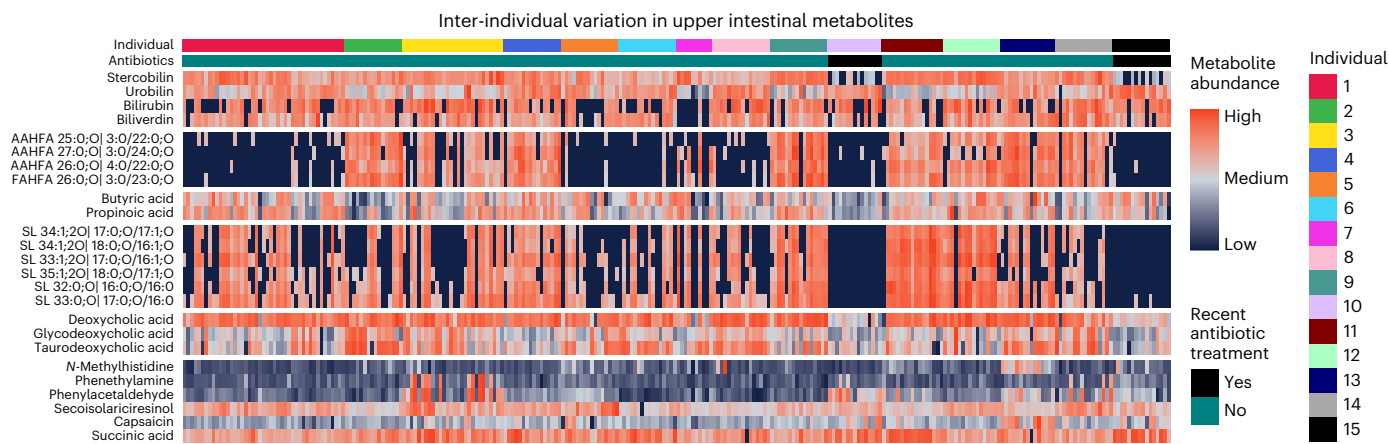
Our dataset exhibited large inter-individual variation (Extended Data Fig. 7). As participants were not prescribed specific meals, diet-based variation was expected to differentiate participants and time points. Using multivariate discriminant analysis (PLS-DA), we identified differences in the proportion of metabolites that were most important for differentiating between the proximal and distal intestine (device types) and among the 15 participants (Extended Data Fig. 6b). The large overall variance among samples obscured clear visualization of PLS-DA based on participants, devices or time points. Nonetheless, the 100 metabolites that contributed most to multivariate discrimination revealed participant-specific trends that were best explained by metabolites of plant and microbial origin (Extended Data Fig. 6b), including the pepper compound capsaicin, the flaxseed compound secoisolariciresinol and the microbially produced butyric acid and propionic acid (Supplementary Table 7). Other metabolites that showed participant-specific variation, such as *N*-methylhistamine, phenethylamine, phenylacetaldehyde and succinic acid, may depend on a combination of human, dietary or microbial factors. Notably, hierarchical clustering separated stool samples by participant, whereas intestinal samples did not strongly cluster by participant (Supplementary Fig. 3), perhaps due to the higher abundance of individual-specific gut microbes in the stool relative to the proximal intestine, which was more dominated by the variability in dietary components.

Although large overall variation obscured direct visualization of inter-individual differences when using all data in PLS-DA projections, specific compounds exhibited very large concentration differences among individuals (Fig. 3). For example, the human heme-derived bile



**Fig. 2 | Intestinal metabolite association with food types.** **a, b, d**, Volcano plots show significance of each metabolite to food types of fruit (**a**), alcohol (**b**) and dessert (**d**) calculated by LMM. Consumption is defined as food eaten within 6 h of swallowing sample devices. Significance of  $P < 0.05$  ( $n = 1,182$  metabolites) is delimited by the lower dashed-dotted horizontal line. Circles indicate non-significance after FDR correction and diamonds indicate 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-dotted lines are  $\pm 0.2$  effect size coefficient. **c**, Chemical enrichment statistics (ChemRICH)

analysis revealed significant chemical classes after fruit consumption visualized by separating classes by chemical lipophilicity ( $\log P$ ) and chemical class significance level of  $-\log_{10}(P)$ . Red circles indicate that the chemical class increased after fruit consumption and blue circle indicates that the chemical class decreased after fruit consumption. Circle size indicates the size of the chemical class. **e**, Theophylline and theobromine levels are strongly associated with caffeine levels. Circles represent measured levels in each sample for which both metabolites were detected. **f**, Chemical diagram of caffeine and known metabolic pathways with structures of detected metabolites and Spearman rank correlation coefficient ( $r_s$ ) for each structure ( $P < 1.0 \times 10^{-13}$  for all metabolites;  $n = 1,182$  metabolites).



**Fig. 3 | Heat map of metabolites with strong inter-participant differences.**

Metabolites include bile pigments, fatty acid esters of hydroxy fatty acids (FAHFAs and AAHFAs), SCFAs, sulfonolipids (SLs) and secondary bile acids. Samples are organized by participant and antibiotic consumption is indicated for the two participants who consumed antibiotics 1 and 5 months before this

study. Colour of heat map ranges from low (blue) to high (red) of metabolite abundance (peak height) or concentration (in nanograms per millilitre) for bile acids. Minimum and maximum values are used to set colour scale for each metabolite (each row).

pigments biliverdin and bilirubin and the microbially produced urobilin and stercobilin varied drastically among devices for some participants and were also greatly reduced or absent in specific participants, such as stercobilin for participants 10 and 15 (Fig. 3). Production of secondary bile acids has been proposed to use a similar enzymatic pathway as stercobilin production<sup>50,51</sup> and the two participants who had low stercobilin levels also showed reduced concentration of deoxycholic acid, a secondary bile acid (Fig. 3). The same participant-specific profiles of bile pigments were also observed in stool samples (Extended Data Fig. 8). Notably, these two participants (10 and 15) were the only ones who reported use of antibiotics in the 6 months before the study. These individuals were characterized by very low levels of stercobilin, deoxycholic acid, a subset of FAHFAs and sulfonolipids (Fig. 3), all of which have been previously linked to the gut microbiota<sup>10,51–53</sup>. Oral antibiotics can affect the intestinal microbiota for more than a year after treatment<sup>54</sup>. These data support the hypothesis that metabolic profiles reflect differences in microbial activity for specific pathways. We did not identify significant associations between microbial species abundance from 16S rRNA gene quantification and stercobilin levels due to limited statistical power, but the presence of FAHFAs and sulfonolipids was associated with the differential abundance of specific microbial species.

FAHFAs were first identified 10 years ago. They are biologically active and regulate physiology<sup>55,56</sup>. Acyl  $\alpha$ -hydroxyl fatty acids (AAHFAs) are a specific subset of FAHFAs with an  $\alpha$ -lipid linkage that were discovered just 2 years ago<sup>53</sup>. We identified 88 FAHFAs in this study (Supplementary Table 1). Notably, four specific FAHFAs exhibited highly consistent differences among participants and all of these FAHFAs are linkages of a long-chain hydroxyl fatty acid backbone esterified with C3- or C4-short-chain moieties. Three of these bioactive lipids are esterified at the  $\alpha$ -position (AAHFA 4:0/22:0, AAHFA 3:0/22:0 and AAHFA 3:0/24:0) and one is esterified elsewhere on the backbone (FAHFA 3:0/23:0). Nine participants frequently exhibited high levels of these FAHFAs, whereas six participants produced very little or undetectable amounts (Fig. 3). As undetectable levels cannot be used for ANOVA statistics, we did not perform significance testing. The same individual-specific trends were observed in stool samples (Extended Data Fig. 8). The only other frequently detected FAHFA with a C3 or C4 sidechain and a long-chain fatty acid was AAHFA 16:0/4:0, which did not follow the same individual-specific trend as the four FAHFAs discussed above. Of note, low levels of long-chain fatty acid or SCFA substrates, propionic (C3:0) and butyric acid (C4:0), did not explain

the observed differences in these short-chain FAHFAs in intestinal or stool samples (Fig. 3 and Extended Data Fig. 9). We therefore investigated whether specific bacteria were associated with the presence or absence of the four FAHFAs of interest. Two taxa, a species in the *Blautia* genus most closely related phylogenetically to *Blautia obeum* and proteobacteria related to *Bilophila wadsworthia*, were significantly associated with detection of these FAHFAs (Supplementary Table 8). *B. obeum* is a known SCFA producer<sup>57</sup>, which suggested that production of the short-chain fatty acyl constituents might be a driver of individual-specific FAHFA production; however, FAHFA 4:0/16:0 did not show the same individual-specific trend, suggesting that the *Blautia* species may specifically produce FAHFAs with propionic and butyric acid esters of hydroxylated very-long-chain fatty acyls (22:0 and 23:0) and not of hydroxyl-forms of the most abundant fatty acids, C16–18. A chemically related group of FAHFAs have been shown to improve glucose homeostasis, stimulate insulin sensitivity and have anti-inflammatory effects<sup>55</sup>. Thus, our findings have potential health implications in addition to providing a link between intestinal chemistry and FAHFA levels in humans.

Here, we report the detection of sulfonolipids in human samples. These lipids were only recently added to lipidomic libraries<sup>58</sup>, leading to the discovery that they are microbially produced in mouse intestinal tracts and linked with both pro- and anti-inflammatory phenotypes<sup>59–61</sup>. In intestinal device samples, we detected sulfonolipids with strong inter-individual trends regardless of sampling time point (meals), suggesting that sulfonolipids were microbially produced in some individuals, but not in others. The two participants who had previously received antibiotics showed much lower levels of sulfonolipids than all other individuals (Fig. 3), suggesting that the antibiotic treatment may have eliminated the microbial producers. Notably, sulfonolipids were also absent in specific samples of other participants, prompting us to examine associations of sulfonolipids with bacterial taxa. Ten taxa were significantly enriched in sulfonolipid-containing samples, including a member of the Desulfosporosporaceae family (Supplementary Table 8), a family that has been associated with ulcerative colitis<sup>62</sup>. Also enriched were two members of Bacteroidetes, a phylum with known sulfonolipid producers<sup>60</sup>. Health implications of sulfonolipid levels in humans are not well understood. Sulfonolipids were recently shown to increase inflammatory responses in macrophages, but decrease inflammation in the presence of lipopolysaccharides<sup>63</sup>. Further investigations are required to determine the relationship of sulfonolipids with human health.

This work provides a report of metabolome differences in the upper intestinal tract in healthy human participants using a non-invasive, ingestible sampling device. Our results 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. An important goal for future investigation is to characterize the effect of antibiotics on intestinal sulfonolipid-, stercobilin- and long-chain AAHFA-producing bacteria and the consequences of such disruptions on health and disease. The disruption of these bacteria by antibiotics may be linked to the incidence and etiology of inflammation, diabetes and inflammatory bowel disease<sup>55,60,63</sup>. Consequently, it will be important to uncover the dynamics and mechanisms of repopulation of antibiotic-treated individuals with these microbes.

In our related companion study<sup>4</sup>, we used the same device samples to broadly study variation in the intestinal environment using several omics approaches. Like the findings on metabolomes reported in this study, we found more pronounced prophage induction in intestinal samples compared to stool and that the host proteome and bile acid profiles varied along the intestines and were highly distinct from those of stool<sup>4</sup>. Complementing the large differences in peptides and amino acids between proximal and distal intestinal locations reported here, in our companion study we show that microbially conjugated bile acid concentrations displayed amino-acid-dependent trends that were not apparent in stool. Taken together, these studies collectively illustrate the utility of sampling directly from the intestines, which should improve our understanding of the intimate relationship between human hosts and their commensal microbes.

As a pilot study, our study has several limitations. First, a larger number of participants will be necessary for making conclusions about the range of gut metabolism within and across human populations and larger studies could also accommodate more detailed microbiome-wide associations of specific bacteria or bacterial families with metabolic conversions. Second, only two participants reported antibiotic use, hence the potential link between antibiotics use and disruption of sulfonolipid and FAHFA metabolism needs to be strengthened in future studies. Third, analysis of temporal variation was limited to two time points per day on two consecutive days, hence changes in metabolic or microbial composition over time were not fully assessed. Fourth, links between upper intestinal microbiota composition and variability to disease conditions, such as small intestinal bacterial overgrowth, will need specific, separate investigations. Despite these limitations, our findings demonstrate that 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 nutritional studies and human disease.

## Methods

### Ingestible sampling device

In a related companion publication<sup>4</sup>, we used the same text to describe the capsule sampling device (CapScan, Envivo Bio). A CapScan consists of hollow elastic collection bladder capped by a one-way valve<sup>4</sup>. To prepare the device for packaging, the collection bladder is evacuated, folded in half and packaged inside a dissolvable capsule measuring 6.5 mm in diameter and 23 mm in length, onto which an enteric coating is applied. The capsule and the enteric coating are designed to prevent contamination of the collection bladder from oral-pharyngeal and gastric microbes during ingestion. The enteric coating and capsule disintegrate when the device reaches the target pH, which is pH 5.5 for type 1, pH 6 for type 2 and pH 7.5 for types 3 and type 4, with type 4 also having a time-delay coating to bias collection toward the ascending colon. After disintegration of the enteric coating, the collection bladder

unfolds and expands into a tube 6 mm in diameter and 33 mm in length, thereby drawing in up to 400  $\mu$ l of gut luminal contents through the one-way valve. The integrity of the sample collected inside the collection bladder is maintained by the one-way valve as the device moves through the colon and is exposed to stool.

In a related companion publication<sup>4</sup>, we used the same text to describe the study design. Each participant concurrently ingested sets of four devices, each with distinct coatings to target the proximal to medial regions of the small intestine (coating types 1 and 2) and more distal regions (coating types 3 and 4). After sampling, the devices were passed in the stool into specimen-collection containers and immediately frozen. After collection, the stool was thawed and devices were retrieved by study staff. For sample removal, the elastic collection bladders were rinsed in 70% isopropyl alcohol and punctured with a sterile hypodermic needle attached to a 1-ml syringe. Samples were transferred into microcentrifuge tubes and pH was measured with an InLab Ultra Micro ISM pH probe (Mettler Toledo). For metabolomics analysis, a 40- $\mu$ l aliquot was spun down for 3 min at 10,000g to collect the supernatant. The rest of the sample was frozen until being thawed for DNA extraction.

### Study design

In a related companion publication<sup>4</sup>, we used the same text to describe the study design. The study was approved by the WIRB-Copernicus Group Institutional Review Board (study no. 1186513) and informed consent was obtained from each participant. Healthy volunteers were selected to exclude participants suffering from clinically detectable gastrointestinal conditions or diseases that would potentially interfere with data acquisition and interpretation.

Participants met all of the following criteria for study inclusion: (1) individuals between the ages of 18 and 70 years; (2) American Society of Anesthesiologists physical status class risk of 1 or 2; (3) for women of childbearing potential, a negative urine pregnancy test within 7 d of screening visit and willingness to use contraception during the entire study period; and (4) fluency in English, understanding the study protocol and able to supply informed written consent, along with complying with study requirements.

Participants with any of the following conditions or characteristics were excluded from the study: (1) history of any of the following: previous gastric or esophageal surgery, including lap banding or bariatric surgery, bowel obstruction, gastric outlet obstruction, diverticulitis, inflammatory bowel disease, ileostomy or colostomy, gastric or esophageal cancer, achalasia, esophageal diverticulum, active dysphagia or odynophagia or active medication use for any gastrointestinal conditions; (2) pregnancy or planned pregnancy within 30 d from screening visit or breast-feeding; (3) any form of active substance abuse or dependence (including drug or alcohol abuse), any unstable medical or psychiatric disorder or any chronic condition that might, in the opinion of the investigator, interfere with conduct of the study; or (4) a clinical condition that, in the judgment of the investigator, could potentially pose a health risk to the individual while involved in the study.

Fifteen healthy individuals were enrolled in this study and each swallowed at least 17 devices over the course of 3 d. Sample size was chosen to assess general variation across human intestinal tracts. Daily instructions included the following guidelines: record all foods and time they were consumed throughout the day; if you work out, do so in the morning; eat breakfast and lunch as usual; swallow a set of four devices 3 h after lunch with up to two-thirds cup of water; do not eat or drink anything for at least 2 h after swallowing devices; if hungry after 2 h, snack lightly (up to 200 calories); do not drink any caffeinated beverages after lunch until the next morning; collect all stool starting 6 h after swallowing this set of devices until 48 h after swallowing the next set of devices; eat dinner as usual at least 6 h after lunch; swallow a set of four CapScan devices 3 h after dinner with two-thirds cup of water; and after swallowing this set, do not eat or drink anything



until the morning. Alcohol consumption and diet contents were not restricted. All ingested devices were recovered and no adverse events were reported during the study. In total, 274 capsule devices provided sufficient material for metabolomics analysis and 225 provided sufficient volume or number of sequencing reads (>2,500) for genomic analysis. Every bowel movement during the study was immediately frozen by the participant at  $-20^{\circ}\text{C}$ . Participant 1 provided additional samples for assessment of replicability. A total of 333 intestinal and stool samples were analysed with metabolomics methods.

### Untargeted metabolomics sample preparation

Sample preparation was performed using a biphasic extraction<sup>64</sup> with water, methanol and methyl tert-butyl ether (MTBE) to separate polar and non-polar metabolites. Capsule device supernatant and stool samples were prepared separately because device samples were liquid and stool samples were solid. For each supernatant sample, 10  $\mu\text{l}$  was aliquoted into one well of a deep sample preparation 96-well plate in a pre-determined randomized order. Samples were extracted one 96-well plate at a time and all steps were carried out at  $4^{\circ}\text{C}$  unless otherwise specified. Between every ten experimental samples, a method blank and external QC sample were prepared. Blanks used 10  $\mu\text{l}$  LC-MS-grade water instead of sample and QC samples used 10  $\mu\text{l}$  pooled sample of human gastrointestinal tract contents from unrelated studies. Then, 170  $\mu\text{l}$  methanol containing SPLASH LIPIDOMIX Mass Spec Standard (Avanti) were added to each well and the plate was heat-sealed with foil, shaken vigorously for 30 s at room temperature, unsealed and 490  $\mu\text{l}$  MTBE was added. The plate was then heat-sealed again, vortexed vigorously for 30 s at room temperature and shaken for 5 min on an orbital shaker. The foil seal was removed and 150  $\mu\text{l}$  LC-MS-grade water was added to each well. The plate was vortexed for 30 s at room temperature and centrifuged at 708g for 12 min. The foil was removed from the deep-well plate and two 180- $\mu\text{l}$  aliquots of the top phase were transferred to two 96-well Vanquish LC plates using a 12-channel pipette. Two 50- $\mu\text{l}$  aliquots of the aqueous phase were then transferred to two other 96-well Vanquish LC plates. All 96-well plates were dried completely under vacuum at room temperature, heat-sealed with foil and stored at  $-80^{\circ}\text{C}$  until further analysis. Each stool sample was prepared by mixing with a spatula and  $5 \pm 1$  mg was transferred to a 2-ml microcentrifuge tube. Then, 225  $\mu\text{l}$  methanol containing SPLASH LIPIDOMIX Mass Spec Standard (Avanti) was added to all microcentrifuge tubes and the tubes were vortexed for 10 s at room temperature. Two 3-mm stainless steel balls and 750  $\mu\text{l}$  MTBE were added to each tube and samples were homogenized in a Geno/Grinder (SPEX) at 1,500 Hz for 1 min. Following that, 188  $\mu\text{l}$  water was added to each tube and each tube was vortexed for 30 s at room temperature. Tubes were centrifuged at 14,000g for 2 min at room temperature. Two aliquots of 180  $\mu\text{l}$  of the organic phase were transferred to two 96-well plates. Two 50- $\mu\text{l}$  aliquots of the aqueous phase were transferred to two 96-well plates. All plates were dried completely in a rotary vacuum evaporator, heat-sealed with foil and stored at  $-80^{\circ}\text{C}$  until further analysis.

### HILIC LC-MS/MS analysis

Samples from the aqueous phase aliquots of sample preparation were removed from  $-80^{\circ}\text{C}$  and 35  $\mu\text{l}$  8:2 acetonitrile:water containing 29 isotopically labeled and synthetic internal standards, including CUDA<sup>65</sup> was added to each well. The 96-well plates were then sealed with foil, vortexed vigorously for 30 s at room temperature, sonicated in a water bath for 5 min at room temperature and centrifuged for 15 min at 708g. Plates were stored in an autosampler at  $4^{\circ}\text{C}$  until analysis (maximum storage time in autosampler was 48 h). LC-MS/MS was performed with a Vanquish LC (Thermo Scientific) coupled to a QExactive HF<sup>+</sup> orbital ion trap mass spectrometer (Thermo Scientific). Chromatographic separation was performed with a Waters Acquity UPLC BEH Amide column (150 mm length  $\times$  2.1 mm inner diameter (i.d.); 1.7- $\mu\text{m}$  particle size) with a 5-mm pre-column (5 mm length  $\times$  2.1 mm i.d.; 1.7- $\mu\text{m}$  particle

size). Mobile phase A was 100% water and B was 95:5 acetonitrile:water. Both mobile phases were modified with 10 mM ammonium formate and 0.125% formic acid. Flow rate was 400  $\mu\text{l min}^{-1}$ , column temperature was  $45^{\circ}\text{C}$  and injection volume was 3  $\mu\text{l}$ . The LC gradient was 100% mobile phase B from 0 to 2 min, 70% B by 7.7 min, 40% B by 9.5 min, 30% by 10.25 min and returned to 100% B by 12.75 min to re-equilibrate until 17 min. All mobile phase gradient shifts were linear. Acetonitrile:water 1:1 was used as a needle wash solvent before and after sample injection. Heated electrospray ionization (HESI) source conditions were as follows: sheath gas flow 50, auxiliary gas flow 13, sweep gas flow 3, capillary temperature  $263^{\circ}\text{C}$ , S-lens RF level 50, auxiliary gas heater temperature  $425^{\circ}\text{C}$  and needle voltage 3,500 V and  $-3,500$  V for positive and negative ionization mode, respectively. Spectra were collected with data-dependent MS/MS acquisition (DDA) for the top four ions. MS scans were collected with 60-k resolving power from 60–900  $m/z$ , AGC target of  $10^6$  ions and maximum accumulation time of 100 ms. MS/MS spectra were collected with 15-k resolving power, 1-Da isolation window, 50-ms maximum accumulation time, a 3-s dynamic exclusion window, stepped (N)CE of 20, 30 and 60 for fragmentation and  $8 \times 10^3$  AGC target. All spectra were stored in centroid mode. Three rounds of iterative exclusion MS/MS were acquired for each pooled QC sample. Immediately after analysis, plates were dried under vacuum, sealed with foil and stored at  $-80^{\circ}\text{C}$ .

### Lipidomics LC-MS/MS analysis

Sample plates from the organic phase aliquots were removed from  $-80^{\circ}\text{C}$  and 50  $\mu\text{l}$  9:1 acetonitrile/toluene was added to each well. The plates were then sealed with foil, vortexed vigorously for 30 s at room temperature, sonicated in a water bath for 5 min at room temperature, centrifuged for 15 min at 708g and transferred to an autosampler kept at  $4^{\circ}\text{C}$  until analysis (maximum storage time in autosampler before analysis was 48 h). Lipidomics LC-MS/MS analysis used a Thermo Scientific Vanquish LC system coupled to a Thermo Scientific QExactive HF<sup>+</sup> orbital ion trap mass spectrometer. Chromatographic separation used a Waters Acquity UPLC CSH C18 column (100 mm in length  $\times$  2.1 mm i.d.; 1.7- $\mu\text{m}$  particle size) with a pre-column (5 mm in length  $\times$  2.1 mm i.d.; 1.7- $\mu\text{m}$  particle size). Mobile phase A was 9:1 acetonitrile:water and B was 8:2 IPA:acetonitrile. For positive-mode electron spray ionization (ESI), the mobile phases were modified with 10 mM ammonium formate and 0.1% formic acid. For negative-mode ESI, mobile phases were modified with 10 mM ammonium acetate. Flow rate was 600  $\mu\text{l min}^{-1}$ , column temperature was  $65^{\circ}\text{C}$  and injection volume was 5  $\mu\text{l}$ . The mobile phase gradient was 15% B from 0 to 0.6 min, 30% B by 2 min, 48% B by 2.5 min, 82% B by 11 min, 99% B from 11.5 to 12 min and 15% B from 12.1 to 14.2 min. HESI source conditions were as follows: sheath gas flow 55, auxiliary gas flow 15, sweep gas flow 3, capillary temperature  $275^{\circ}\text{C}$ , S-lens RF level 50, auxiliary gas heater temperature  $450^{\circ}\text{C}$  and needle voltage 3,500 V and  $-3,500$  V for positive and negative ionization mode, respectively. DDA MS/MS spectra were acquired for the top four ions. MS scans were collected with 60-k resolving power from 120–1,700  $m/z$ , AGC target of  $10^6$  ions and maximum accumulation time of 100 ms. MS/MS spectra were collected with 15-k resolving power, 1-Da isolation window, normalized collision energy of 20, 30 and 60, 2-s dynamic exclusion window,  $8 \times 10^3$  AGC target and 50-ms maximum accumulation time. Spectra were stored in centroid mode. Three rounds of iterative exclusion MS/MS were acquired for each pooled QC sample. Immediately after all samples were analysed, the plates were dried under vacuum, sealed with foil and stored at  $-80^{\circ}\text{C}$ .

### Untargeted GC-MS analysis

Dried aqueous-phase samples as described above were removed from  $-80^{\circ}\text{C}$  and 10  $\mu\text{l}$  40  $\text{mg ml}^{-1}$  methoxyamine hydrochloride in pyridine was added to each well. The plates were sealed and shaken at  $30^{\circ}\text{C}$  for 90 min. Foil was removed and 90  $\mu\text{l}$  *N*-methyl-*N*-trimethylsilyl trifluoroacetamide (containing internal standards of C8-C30 fatty

acid methyl esters) was added to each well and plates were shaken for 30 min at 37 °C. The plate was then centrifuged at 2,400g for 15 min at room temperature. Foil was removed and 90 µl supernatant was transferred to glass crimp-top vials with glass insert and crimped shut. Samples were analysed within 48 h of derivatization. GC–MS analysis was carried out as previously described<sup>66</sup>. Briefly, 0.5 µl was analysed in splitless mode through an Agilent 6890 GC equipped with a RTX-5Sil MS column (30 m × 0.25 mm i.d., 0.25 µm 95:5 dimethyl diphenyl polysiloxane film). Chromatography used helium at 1 ml min<sup>-1</sup> and temperature gradient of 50 °C to 275 °C. MS analysis was performed using a Leco Pegasus III TOF mass spectrometer. Spectra were collected at 17 spectra per second with 70 eV EI. Data files were pre-processed in Leco ChromaTOF software and further analysed using the automated GC–MS data-processing pipeline BinBase<sup>67</sup> with FiehnLib for spectra and retention time-matching.

### Bile acid quantification

Bile acids were targeted and quantified as previously reported<sup>4</sup>. Briefly, dried aqueous-phase sample plates from hydrophilic interaction chromatography (HILIC) ESI+ analysis were removed from a –80 °C freezer and dissolved in 60 µl bile acid run solvent (1:1 ACN methanol containing 100 ng ml<sup>-1</sup> of five isotopically labeled bile acid internal standards). Plates were vortexed for 30 s and sonicated in a water bath for 5 min. All samples were diluted 30-fold in bile acid run solvent. All plates were centrifuged at 708g for 15 min at 4 °C and stored in a Vanquish autosampler at 4 °C until analysis. A nine-point standard curve was prepared between 0 to 1,500 ng ml<sup>-1</sup> and analysed with samples. A Vanquish UPLC and Altis QqQ mass spectrometer (Thermo Fisher Scientific) were used for targeted LC–MS/MS analysis. An Acquity BEH C18 column (100 mm × 2.1 mm i.d., 1.7-µm particle size; Waters) was used with mobile phase A of water and B of acetonitrile, both with 0.1% formic acid. Skyline software and custom R scripts were used to process and quantify bile acid analytes.

### Short-chain fatty acid quantification

SCFA quantification was performed through derivatization with *N*-tert-butyltrimethylsilyl-*N*-methyltrifluoroacetamide (MTBSTFA) based on a previously reported protocol<sup>68</sup>. Briefly, 10 µl intestinal supernatant was transferred to a 1.5-ml microcentrifuge tube on ice pre-filled with 50 µl LC–MS-grade water containing deuterium-labeled SCFA internal standards and 10 µl 37% hydrochloric acid. Then, 100 µl MTBE also containing deuterium-labeled SCFA internal standards was added to each tube and the tubes were shaken on rotary shaker plate for 30 min at room temperature. Tubes were centrifuged at 14,000g for 2 min and 20 µl MTBE (top layer) was transferred to a crimp-top GC–MS vial fitted with low recovery insert. Then, 5 µl MTBSTFA was added to each vial, which was then sealed. Vials were shaken on an orbital shaker plate for 30 min at 80 °C, cooled to room temperature and analysed by GC–MS. An Agilent 6890 GC coupled to a Leco Pegasus IV TOF mass spectrometer was used with a DB5 DuraGuard (30 m × 0.25 mm × 0.25 µm) capillary column. Following that, 1 µl was injected with 1:10 split mode enabled, helium flow rate of 1.2 ml min<sup>-1</sup>, temperature ramped from 50 °C to 290 °C over 20.8 min and scan rate of 17 spectra per min from 50–550 *m/z*. A six-point standard curve between 1 and 100 µg ml<sup>-1</sup> was analysed for every 80 samples and blank samples were analysed between every ten experimental samples. Samples and standard curve mixes were prepared within 24 h of GC–MS analysis. Data were converted to Agilent (.d) format and processed using Agilent MassHunter Quant v.B.09.00. Linear six-point standard curves using ratio of analyte to internal standard were used to quantify SCFAs in experimental samples.

### Untargeted LC–MS/MS data processing

MS-DIAL v.4.80 (ref. 58) was used to process untargeted LC–MS/MS data. HILIC LC–MS/MS (ESI+ and ESI–) and lipidomics (ESI+ and ESI–)

datasets were processed with manually optimized parameters (Supplementary Table 9). Peak height was reported for all untargeted analyses. Experimental MS/MS spectra were matched to MassBank of North America as well as NIST20 spectral libraries for HILIC analyses and all lipid reference MS/MS spectra from MS-DIAL were used for lipidomics analyses. Retention time information for HILIC and lipidomics from authentic standards run under identical chromatography conditions was used as another line of evidence for metabolite annotation. Metabolite annotations were based on accurate mass matching with retention time and/or an MS/MS match to a library spectrum. Blank subtraction was performed by retaining LC–MS features for which the maximum intensity from an experimental sample was at least ten times as high as the average of the method blanks, along with further data reduction as described in Supplementary Table 2. The final step of data processing was manual investigation of annotated features for MS/MS match quality, peak quality, peak alignment and removal of in-source fragments using correlation between features with close retention time for in-source fragment identification. When multiple metabolites from spectral libraries matched one experimental MS/MS, the match was recorded as a non-unique MS/MS match (Supplementary Table 1). Predicted retention times calculated using Retip<sup>69</sup> were used as an additional line of evidence to flag low-quality annotations.

### DNA extraction and 16S rRNA gene sequence analysis

DNA was extracted using a Microbial DNA extraction kit (QIAGEN)<sup>70</sup> and 50 µl from a capsule device or 100 mg stool. The 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 300-bp reads on a MiSeq (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 (ref. 71) and DADA2 as previously described using pseudo-pooling<sup>72</sup>. *truncLenF* and *truncLenR* parameters were set to 250 and 180, respectively. Taxonomy was assigned using the Silva rRNA database v.132 (ref. 73). Samples with >2,500 reads were retained for analyses.

### Statistics and data analysis

Statistical tests were performed using R<sup>71</sup>. LMMs were performed using the *lmerTest* and *lme4* R packages. To examine spatial differences across the intestine, we applied LMMs that accounted for sampling location (proximal (device types 1 and 2) or distal (device types 3 and 4)), eight food types classified from food logs (vegetable and animal protein (meat, egg and fish), grain (rice, pasta, bread and other grain), coffee or tea, dessert or alcohol (beer, wine or alcoholic seltzer), dairy and fruit) and antibiotic consumption within 6 months as fixed effect variables and inter-individual variation as a random effect variable (Supplementary Tables 3 and 4). Food types were manually assigned from participants' written food logs using customized assessment forms. Metabolite abundances were log<sub>10</sub>-transformed and missing values were treated as zeros, followed by scaling from –1 and 1 before LMM analysis. A Benjamini–Hochberg<sup>74</sup> correction was used to account for multiple-hypothesis testing. Correlations between the microbiota (log<sub>2</sub>-scaled amplicon sequence variant abundance) and metabolites (log<sub>10</sub>-scaled metabolite abundance) were calculated at the amplicon sequence variant-level using Pearson's correlations. Differential abundance analysis using Limma-Voom differential abundance was utilized for FAHFs and sulfonolipids in a presence/absence analysis. Only taxa with an absolute differential abundance

>0.75 and Benjamini–Hochberg-corrected  $P < 0.1$  were considered. ChemRICH<sup>75</sup> was used to calculate enrichment statistics. Clustering was performed using the `hclust` function with the metabolite Spearman rank correlation matrix calculated using the `cor` function in R and Euclidean distance calculated with the `as.dist` function in R. PLS-DA and principal-component analysis (PCA) were performed with the `ropls` package in R<sup>76</sup>. PLS-DA models to distinguish participant and device type were assessed by sevenfold cross validation. Using 20–1,000 random permutations of class labels performed by the `ropls` R package to test for overfitting, models maintained  $Q^2Y > 0.15$  and  $P < 0.05$  (ref. 77). Untargeted LC–MS/MS (HILIC and RP ESI+/-) features were normalized to the sum of internal standards for each platform, which has been shown to be more robust than normalizations to single compounds<sup>78</sup>. This normalization was performed by dividing each LC–MS feature by the sum of internal standard peak heights for that sample<sup>78,79</sup>. GC–MS data were normalized to the summed intensity of all annotated metabolites as extensively discussed in published protocols<sup>80</sup>. This method addresses differences specific to GC methods, recently called normalization to the total useful peak area<sup>81</sup>. Pooled QC data were found in a dense cluster when compared to CapScan and stool samples (Extended Data Fig. 1). During merging of datasets, metabolites detected by multiple assays were simplified to keep only data from one instrument, with preference for retaining data from the assay with lower technical variance (% coefficient of variance of pooled QC). Metabolites that were detected only in a single assay remained in the dataset, independent of the % coefficient of variance of pooled QC (Supplementary Table 1).  $\log_{10}$  transformation and zero-value imputation using one-tenth of the minimum reported peak height for non-detected features was performed for each metabolite before PCA and PLS-DA.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Raw mass spectrometry data are available on the Metabolomics Workbench (<https://www.metabolomicsworkbench.org/>) under studies ST002073, ST002075, ST002407, ST002409 and ST002411. The 16S and metagenomics sequencing reads are available on NCBI under BioProject PRJNA822660. Taxonomy was assigned using the Silva rRNA database v.132 (<https://www.arb-silva.de/>). Mass spectra were annotated using MassBank of North America public libraries (<https://massbank.us/>) and NIST20 libraries licensed from NIST. Scripts are archived at Zenodo ([https://zenodo.org/record/7659119#.ZBGhMh\\_P23A](https://zenodo.org/record/7659119#.ZBGhMh_P23A)).

### References

- Bloszies, C. S. & Fiehn, O. Using untargeted metabolomics for detecting exposome compounds. *Curr. Opin. Toxicol.* **8**, 87–92 (2018).
- Folz, J., Shalon, D. & Fiehn, O. Metabolomics analysis of time-series human small intestine lumen samples collected in vivo. *Food Funct.* **12**, 9405–9415 (2021).
- Hoffmann, M. A. et al. High-confidence structural annotation of metabolites absent from spectral libraries. *Nat. Biotechnol.* **40**, 411–421 (2022).
- Shalon, D. et al. Profiling of the human intestinal microbiome and bile acids under physiologic conditions using an ingestible sampling device. *Nature* <https://doi.org/10.1038/s41586-023-05989-7> (2023).
- Evans, D. F. et al. Measurement of gastrointestinal pH profiles in normal ambulant human subjects. *Gut* **29**, 1035–1041 (1988).
- Schymanski, E. L. et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ. Sci. Technol.* **48**, 2097–2098 (2014).
- Djombou Feunang, Y. et al. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* **8**, 1–20 (2016).
- Ríos-Covián, D. et al. Intestinal short chain fatty acids and their link with diet and human health. *Front. Microbiol.* **7**, 1–9 (2016).
- 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 (1995).
- Ridlon, J. M., Kang, D.-J. & Hylemon, P. B. Bile salt biotransformations by human intestinal bacteria. *J. Lipid Res.* **47**, 241–259 (2006).
- Lucas, L. N. et al. Dominant bacterial phyla from the human gut show widespread ability to transform and conjugate bile acids. *mSystems* **6**, e00805–e00821 (2021).
- Quinn, R. A. et al. Global chemical impact of the microbiome includes novel bile acid conjugations. *Nature* **579**, 123–129 (2020).
- 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 (2003).
- Bhutipia, Y. D. & Ganapathy, V. in *Physiology of the Gastrointestinal Tract*, Ch. 47 (Elsevier, 2018).
- 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 (1996).
- Reuter, S. E. & Evans, A. M. Carnitine and acylcarnitines: pharmacokinetic, pharmacological and clinical aspects. *Clin. Pharmacokinet.* **51**, 553–572 (2012).
- Nilsson, Å. & Duan, R. D. Alkaline sphingomyelinases and ceramidases of the gastrointestinal tract. *Chem. Phys. Lipids* **102**, 97–105 (1999).
- Zhang, X. et al. Widespread protein lysine acetylation in gut microbiome and its alterations in patients with Crohn's disease. *Nat. Commun.* **11**, 1–12 (2020).
- Arnaud, A., Ramírez, M., Baxter, J. H. & Angulo, A. J. Absorption of enterally administered N-acetyl-L-glutamine versus glutamine in pigs. *Clin. Nutr.* **23**, 1303–1312 (2004).
- Stegink, L. D., Filer, L. J. & 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 (1980).
- Eran Nagar, E., Okun, Z. & Shpigelman, A. Digestive fate of polyphenols: updated view of the influence of chemical structure and the presence of cell wall material. *Curr. Opin. Food Sci.* **31**, 38–46 (2020).
- 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 (2000).
- Parada, J. & Aguilera, J. M. Food microstructure affects the bioavailability of several nutrients. *J. Food Sci.* **72**, 21–32 (2007).
- 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 (2010).
- 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 (2010).
- Kuijsten, A., Arts, I. C. W., Van't Veer, P. & Hollman, P. C. H. The relative bioavailability of enterolignans in humans is enhanced by milling and crushing of flaxseed. *J. Nutr.* **135**, 2812–2816 (2005).
- Jenab, M. & Thompson, L. U. The influence of flaxseed and lignans on colon carcinogenesis and  $\beta$ -glucuronidase activity. *Carcinogenesis* **17**, 1343–1348 (1996).

28. Wanders, R. J. A., Komen, J. & Kemp, S. Fatty acid omega-oxidation as a rescue pathway for fatty acid oxidation disorders. *FEBS J.* **278**, 182–194 (2011).
29. Miura, Y. The biological significance of  $\omega$ -oxidation of fatty acids. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **89**, 370–382 (2013).
30. 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 (2003).
31. McKenna, E. J. & Coon, M. J. Enzymatic  $\omega$ -oxidation. *J. Biol. Chem.* **245**, 3882–3889 (1970).
32. 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 (2000).
33. 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 (2014).
34. Høiseth, G., Morini, L., Poletini, A., Christophersen, A. & Mørland, J. Blood kinetics of ethyl glucuronide and ethyl sulphate in heavy drinkers during alcohol detoxification. *Forensic Sci. Int.* **188**, 52–56 (2009).
35. Servillo, L. et al. Betaines in fruits of citrus genus plants. *J. Agric. Food Chem.* **59**, 9410–9416 (2011).
36. Hossain, G. S. et al. L-Amino acid oxidases from microbial sources: types, properties, functions, and applications. *Appl. Microbiol. Biotechnol.* **98**, 1507–1515 (2014).
37. Deng, X., Wood, P. G., Eagon, P. K. & Whitcomb, D. C. Chronic alcohol-induced alterations in the pancreatic secretory control mechanisms. *Digest. Dis. Sci.* **49**, 805–819 (2004).
38. Judák, L. et al. Ethanol and its non-oxidative metabolites profoundly inhibit CFTR function in pancreatic epithelial cells which is prevented by ATP supplementation. *Pflug. Arch. Eur. J. Physiol.* **466**, 549–562 (2014).
39. Kotormán, M., Laczkó, I., Szabó, A. & Simon, L. M. Effects of  $\text{Ca}^{2+}$  on catalytic activity and conformation of trypsin and  $\alpha$ -chymotrypsin in aqueous ethanol. *Biochem. Biophys. Res. Commun.* **304**, 18–21 (2003).
40. Panoutopoulos, 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 (2005).
41. 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* **2013**, 1–8 (2013).
42. 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 (2013).
43. 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 (2013).
44. Blanchard, J. & Sawers, S. J. A. The absolute bioavailability of caffeine in man. *Eur. J. Clin. Pharmacol.* **24**, 93–98 (1983).
45. Tassaneeyakul, W. et al. Caffeine metabolism by human hepatic cytochromes p450: contributions of 1A2, 2E1 and 3A isoforms. *Biochem. Pharmacol.* **47**, 1767–1776 (1994).
46. Gummadi, S. N., Bhavya, B. & Ashok, N. Physiology, biochemistry and possible applications of microbial caffeine degradation. *Appl. Microbiol. Biotechnol.* **93**, 545–554 (2012).
47. 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 (2015).
48. Arnaud, M. J. in *Caffeine, Coffee and Health* (ed Garattini, S.) 43–95 (Raven Press, 1993).
49. 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).
50. Hamoud, A. R., Weaver, L., Stec, D. E. & Hinds, T. D. Bilirubin in the liver–gut signaling axis. *Trends Endocrinol. Metab.* **29**, 140–150 (2018).
51. Vitek, L. & Ostrow, J. Bilirubin chemistry and metabolism; harmful and protective aspects. *Curr. Pharm. Des.* **15**, 2869–2883 (2009).
52. Saxerholt, H. et al. Influence of antibiotics on the faecal excretion of bile pigments in healthy subjects. *Scand. J. Gastroenterol.* **21**, 991–996 (1986).
53. Yasuda, S. et al. Elucidation of gut microbiota-associated lipids using LC–MS/MS and 16S rRNA sequence analyses. *iScience* **23**, 101841 (2020).
54. Zimmermann, P. & Curtis, N. The effect of antibiotics on the composition of the intestinal microbiota - a systematic review. *J. Infect.* **79**, 471–489 (2019).
55. Yore, M. M. et al. Discovery of a class of endogenous mammalian lipids with anti-diabetic and anti-inflammatory effects. *Cell* **159**, 318–332 (2014).
56. 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 (2018).
57. Xu, Y., Zhu, Y., Li, X. & Sun, B. Dynamic balancing of intestinal short-chain fatty acids: the crucial role of bacterial metabolism. *Trends Food Sci. Technol.* **100**, 118–130 (2020).
58. Tsugawa, H. et al. A lipidome atlas in MS-DIAL 4. *Nat. Biotechnol.* **38**, 1159–1163 (2020).
59. Gowda, S. G. B. et al. Identification of short-chain fatty acid esters of hydroxy fatty acids (SFAHFAs) in a murine model by nontargeted analysis using ultra-high-performance liquid chromatography/linear ion trap quadrupole-Orbitrap mass spectrometry. *Rapid Commun. Mass Spectrom.* **34**, e8831 (2020).
60. Walker, A. et al. Sulfonolipids as novel metabolite markers of *Alistipes* and *Odoribacter* affected by high-fat diets. *Sci. Rep.* **7**, 1–10 (2017).
61. Maeda, J. et al. Inhibitory effects of sulfobacin B on DNA polymerase and inflammation. *Int. J. Mol. Med.* **26**, 751–758 (2010).
62. Gibson, G. R., Cummings, J. H. & Macfarlane, G. T. Growth and activities of sulphate-reducing bacteria in gut contents of healthy subjects and patients with ulcerative colitis. *FEMS Microbiol. Lett.* **86**, 103–112 (1991).
63. Hou, L. et al. Identification and biosynthesis of pro-inflammatory sulfonolipids from an opportunistic pathogen *Chryseobacterium gleum*. *ACS Chem. Biol.* **17**, 1197–1206 (2022).
64. 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 (2008).
65. Barupal, D. K. et al. A comprehensive plasma metabolomics dataset for a cohort of mouse knockouts within the international mouse phenotyping consortium. *Metabolites* **9**, 9050101 (2019).
66. Beaudoin, G. A. W. et al. Salvage of the 5-deoxyribose byproduct of radical SAM enzymes. *Nat. Commun.* **9**, 3105 (2018).
67. Skogerson, K., Wohlgemuth, G., Barupal, D. K. & Fiehn, O. The volatile compound BinBase mass spectral database. *BMC Bioinform.* **12**, 321 (2011).
68. Nusbaum, D. J. et al. Gut microbial and metabolomic profiles after fecal microbiota transplantation in pediatric ulcerative colitis patients. *FEMS Microbiol. Ecol.* **94**, 1–11 (2018).
69. Bonini, P., Kind, T., Tsugawa, H., Barupal, D. K. & Fiehn, O. Retip: retention time prediction for compound annotation in untargeted metabolomics. *Anal. Chem.* **92**, 7515–7522 (2020).

70. Celis, A. I. et al. Optimization of the 16S rRNA sequencing analysis pipeline for studying in vitro communities of gut commensals. *iScience* **25**, 103907 (2022).
  71. R Foundation for Statistical Computing. R: a language and environment for statistical computing. *R Foundation* <http://www.R-project.org/> (2016).
  72. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
  73. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, 590–596 (2013).
  74. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
  75. Barupal, D. K. & Fiehn, O. Chemical similarity enrichment analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets. *Sci. Rep.* **7**, 14567 (2017).
  76. Thévenot, E. A., Roux, A., Xu, Y., Ezan, E. & Junot, C. Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *J. Proteome Res.* **14**, 3322–3335 (2015).
  77. Szymańska, E., Saccenti, E., Smilde, A. K. & Westerhuis, J. A. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **8**, 3–16 (2012).
  78. Sysi-Aho, M., Katajamaa, M., Yetukuri, L. & Orešič, M. Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC Bioinform.* **8**, 93 (2007).
  79. De Livera, A. M. et al. Normalizing and integrating metabolomics data. *Anal. Chem.* **84**, 10768–10776 (2012).
  80. Fiehn, O. Metabolomics by gas chromatography–mass spectrometry: combined targeted and untargeted profiling. *Curr. Protoc. Mol. Biol.* **114**, 30–34 (2016).
  81. Nam, S. L., de la Mata, A. P., Dias, R. P. & Harynuk, J. J. Towards standardization of data normalization strategies to improve urinary metabolomics studies by GC× GC-TOFMS. *Metabolites* **10**, 376 (2020).
- statistical analyses and produced figures and tables. G.T. recruited participants and obtained consents and samples. K.C.H., R.N.C., J.G. and D.A.R. acquired, processed and analysed microbiome data. J.F. and O.F. wrote the manuscript with contributions from D.S., K.C.H., G.T. and R.N.C. All authors approved the final version of the manuscript.

### Competing interests

D.S. is an employee and has equity interest in Envivo Bio, a company with interest in the human gastrointestinal tract. All other authors declare no competing interests.

### Ethics approval

The study was reviewed and approved by an independent organization, the Copernicus Group within the Western Institutional Review Board under study no. 1186513. Informed consent was obtained from each participant.

### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s42255-023-00777-z>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42255-023-00777-z>.

**Correspondence and requests for materials** should be addressed to Oliver Fiehn.

**Peer review information** *Nature Metabolism* thanks Robert Quinn and the other, anonymous, reviewers for their contribution to the peer review of this work. Primary Handling Editor: Ashley Castellanos-Jankiewicz, in collaboration with the *Nature Metabolism* team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

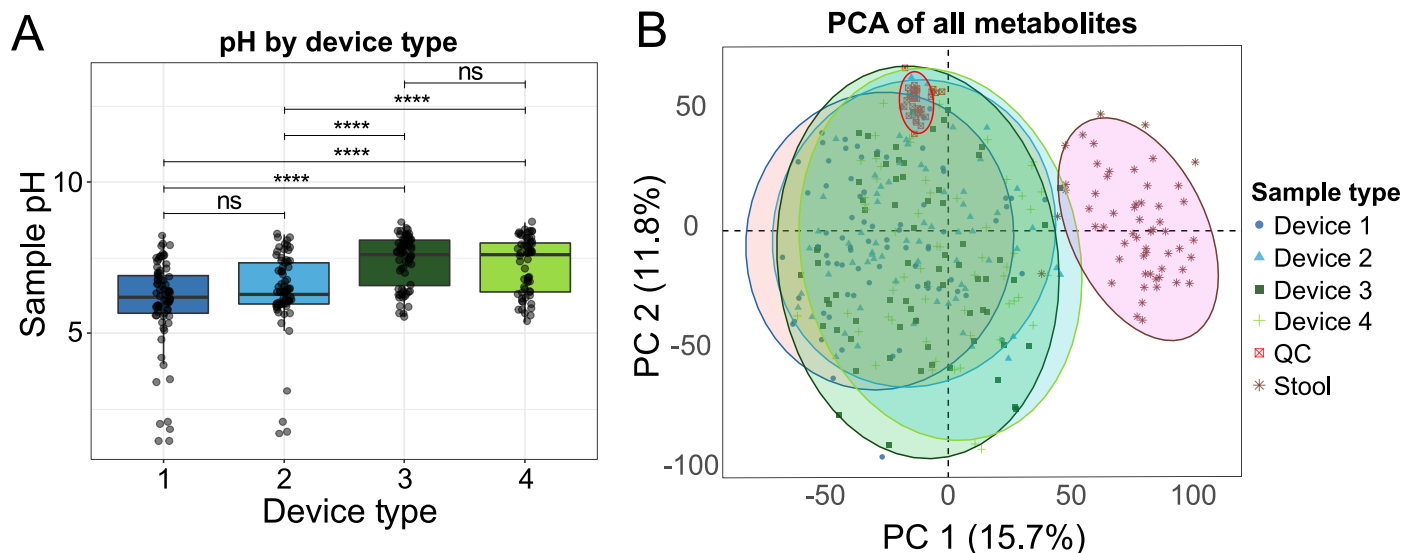
© The Author(s) 2023

### Acknowledgements

J. Wells and Z. Tippins assisted in mass spectrometry analyses at the West Coast Metabolomics Center. R code from D. Barupal was used as downloaded from his GitHub site for ChemRICH assessments. C. Brydges consulted on statistical regression analyses. This study was funded by USDA 2021-67017-35783 to O.F., National Institutes of Health (NIH) grant RM1 GM135102 to K.C.H. and National Science Foundation grant EF-2125383 to K.C.H. and salary for J.F. 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 D.S. J.G. was supported by a Stanford University School of Medicine Dean's Fellowship and by NIH F32 DK130574. D.A.R. is supported by the Thomas C. and Joan M. Merigan Endowment at Stanford University. K.C.H. and D.A.R. are Chan Zuckerberg Biohub Investigators.

### Author contributions

O.F., K.C.H. and D.S. designed the study. J.F. acquired and processed metabolome data, assisted by J.M.M. J.F. performed

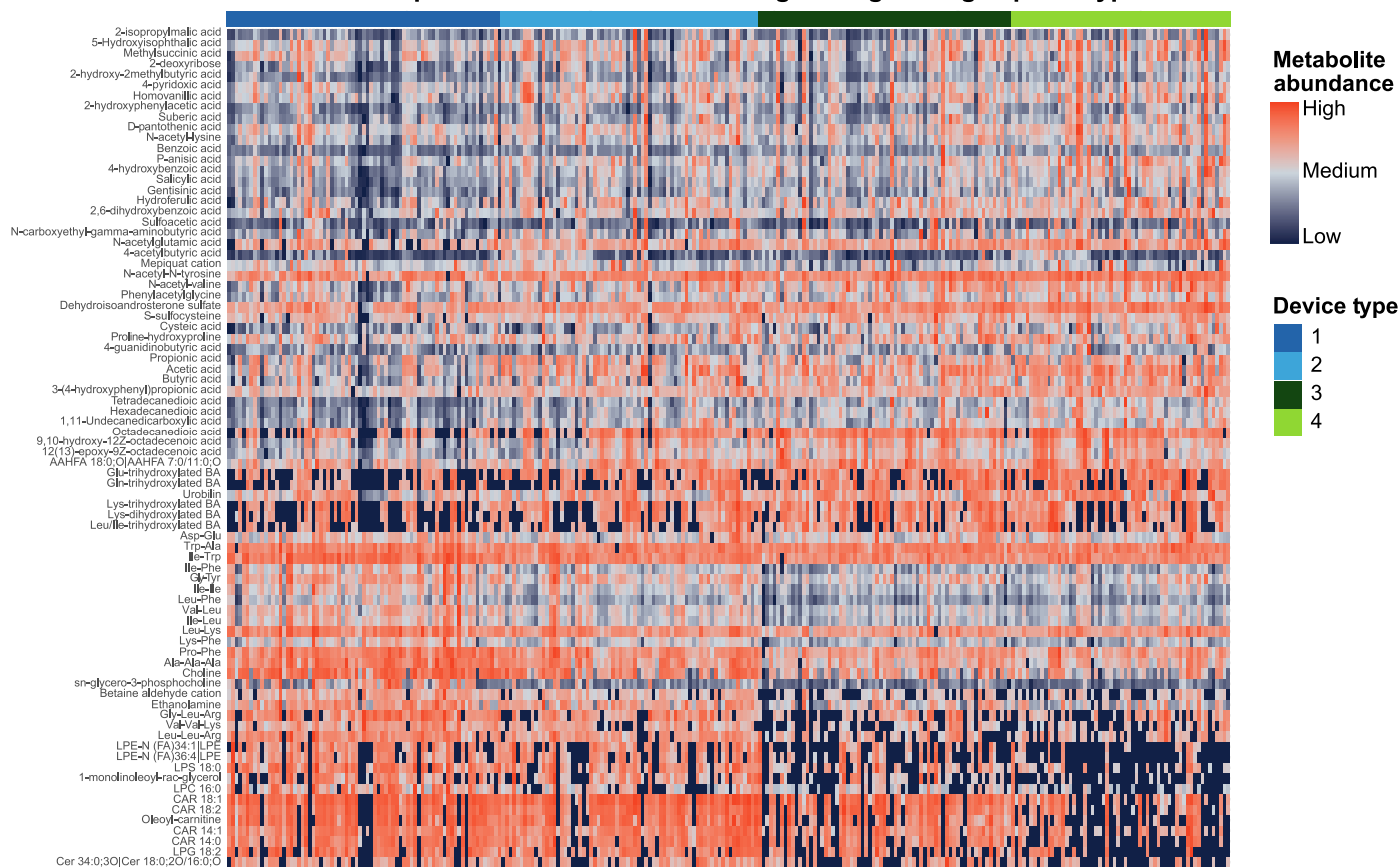


**Extended Data Fig. 1 | Summary visualizations.** **a**, Sample pH for intestinal sampling device types. Boxes represent  $\pm 1$  interquartile range, and the line extends to the sample furthest from the median sample with a maximum extension of 1.5 interquartile ranges away from the median. All samples are plotted as individual points. (n = 274 total devices, Device type 1 = 75, Device

type 2 = 70, Device type 3 = 69, Device type 4 = 60) \*\*\*\*:  $p < 0.0001$  (Wilcoxon signed-rank). **b**, Principal Component Analysis (PCA) of normalized peak heights from the metabolomics dataset. QC samples are pooled intestinal contents from human intestinal tract samples.



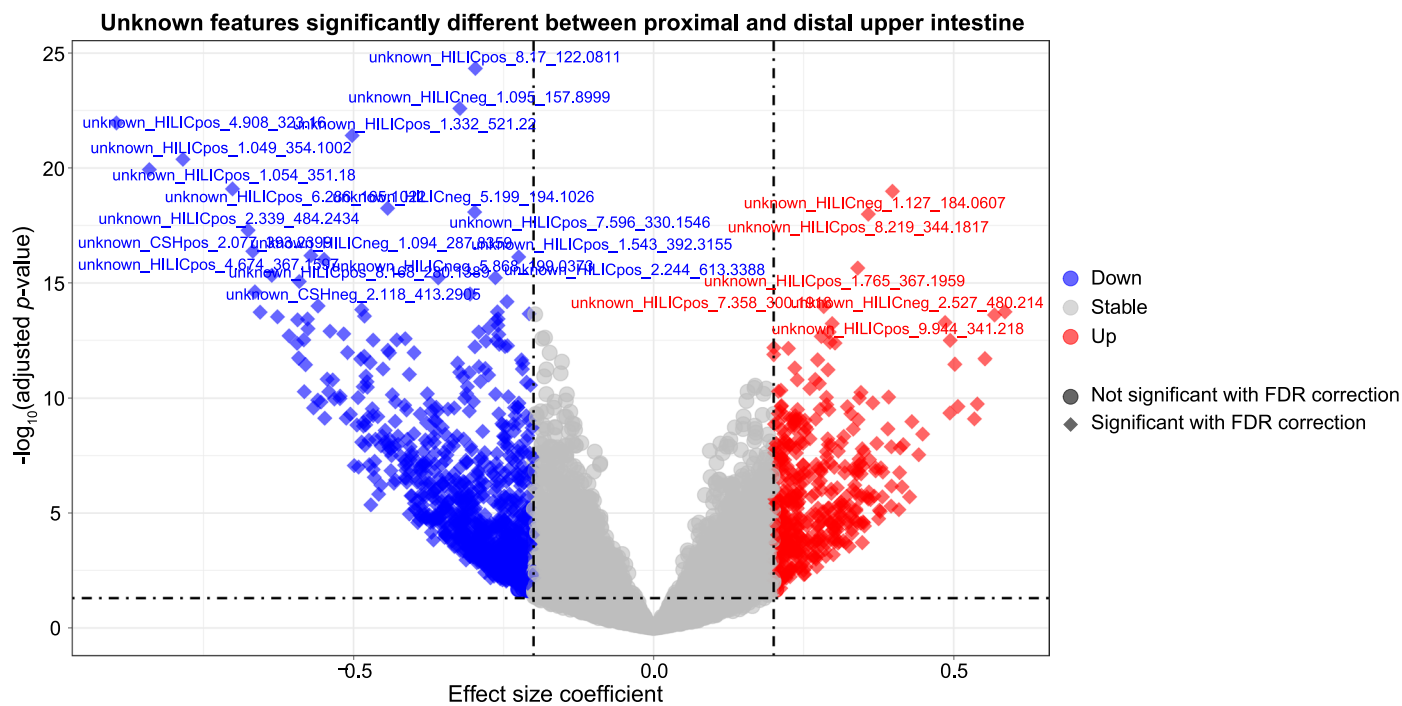
## Most important metabolites for distinguishing among capsule types



**Extended Data Fig. 3 | The top 80 metabolites most important for distinguishing device samples by type as determined by partial least square discriminant analysis (PLS-DA). Samples are organized on the x-axis by device**

type, and metabolites are organized on the y-axis according to hierarchical clustering. Color represents metabolite abundance, and the minimum and maximum values for color scale are set individually for each metabolite (row).

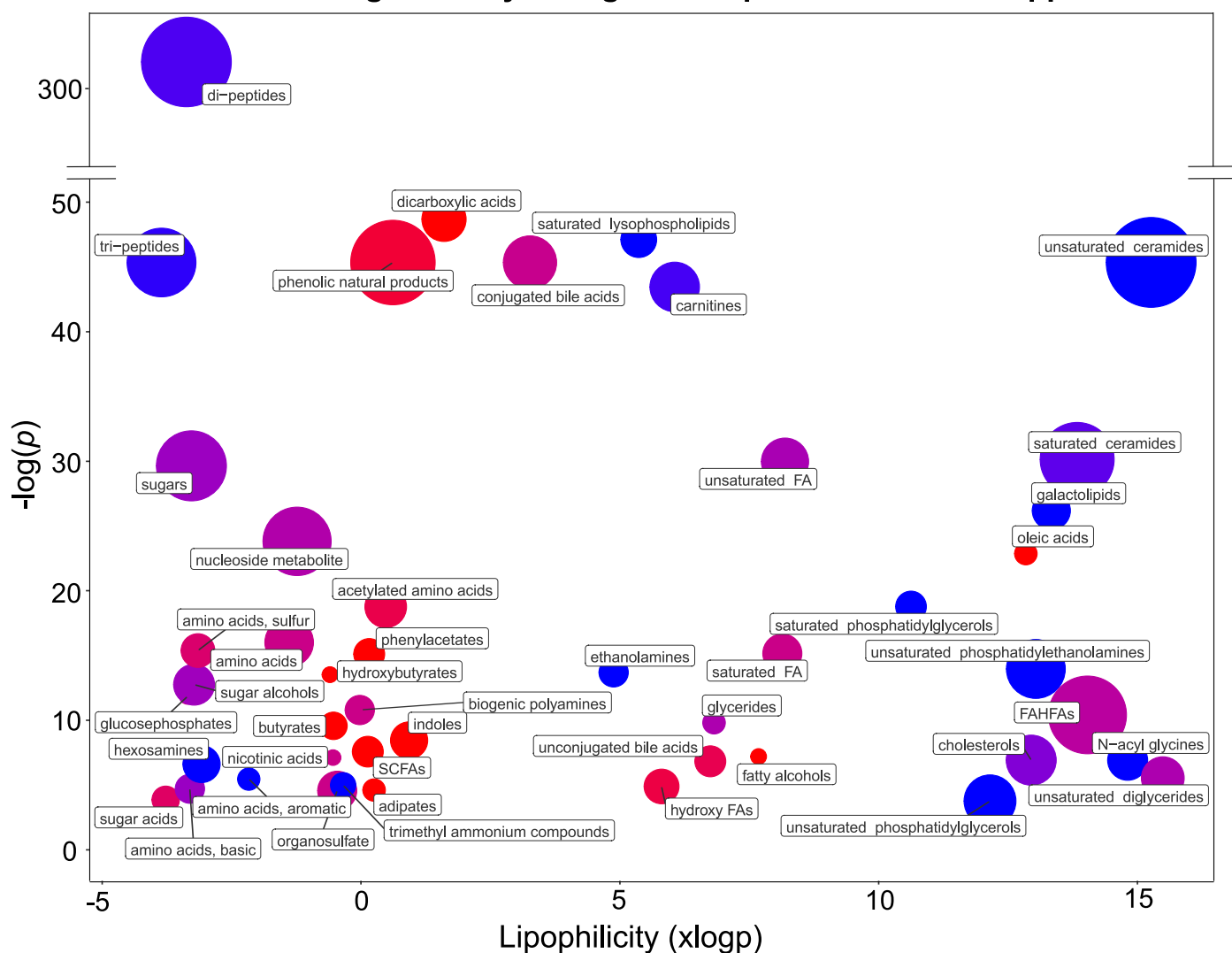




**Extended Data Fig. 4 | Comparison of levels of unannotated features in the proximal and distal upper intestines.** Data from devices 1+2 (proximal) compared to data from devices 3+4 (distal). Significance was calculated using a linear mixed effect model (LMM). The horizontal dashed line represents the significance threshold of  $p < 0.05$  (For exact  $p$  values see Supplementary Table 4). Circles indicate metabolites with non-significant differences after false discovery rate (FDR) correction or an effect size coefficient  $< \pm 0.2$ . Diamonds indicate

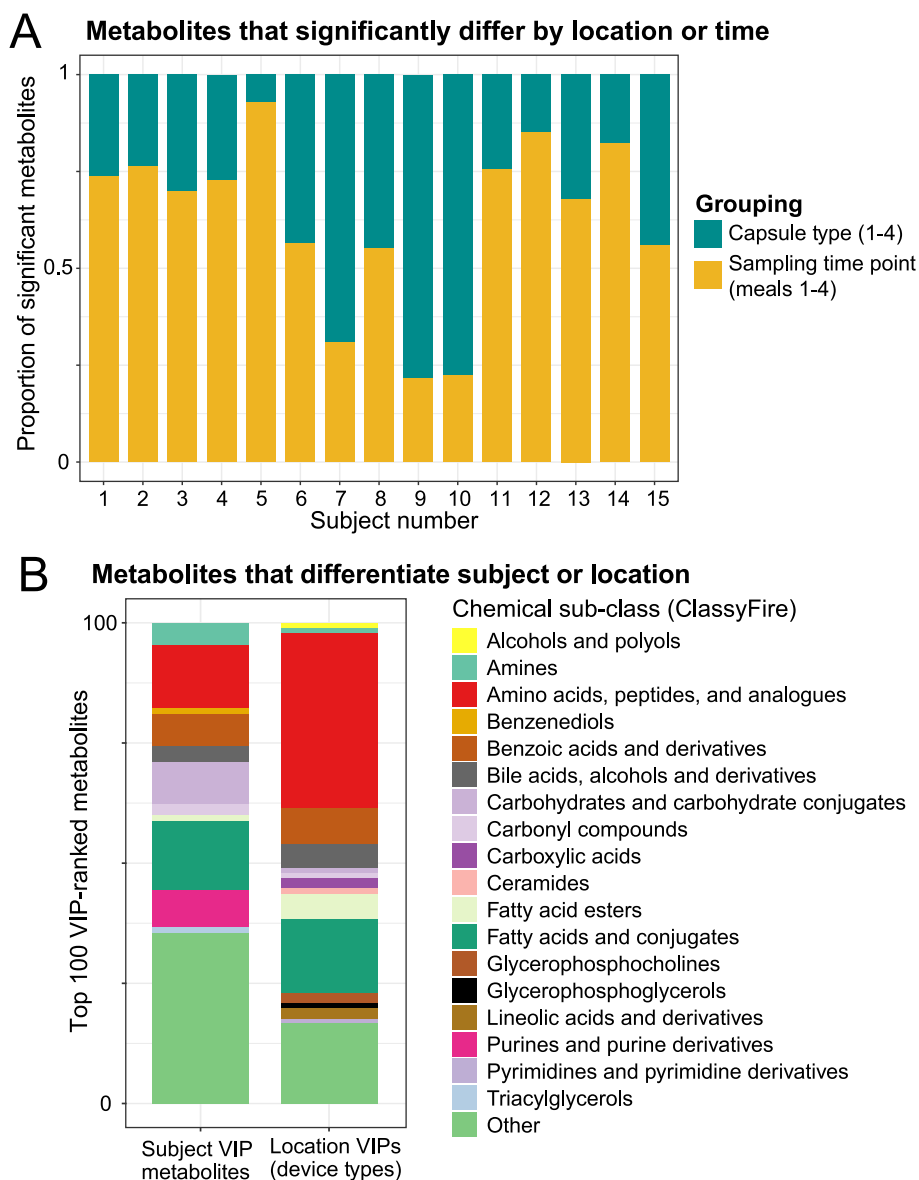
metabolites with significant ( $p < 0.05$ ) differences after FDR correction ( $n = 9,317$ ) and effect size coefficient  $> \pm 0.2$ . Only features detected in  $>50\%$  of intestinal samples were included in this analysis ( $n = 9,317$  features). Effect size coefficient is the slope estimated by the LMM, with positive (negative) coefficient representing a metabolite that is higher (lower) in the distal compared to proximal upper intestine. Vertical dashed lines are  $\pm 0.20$  times the effect size coefficient.

## Chemical clusters significantly changed from proximal to distal upper intestine



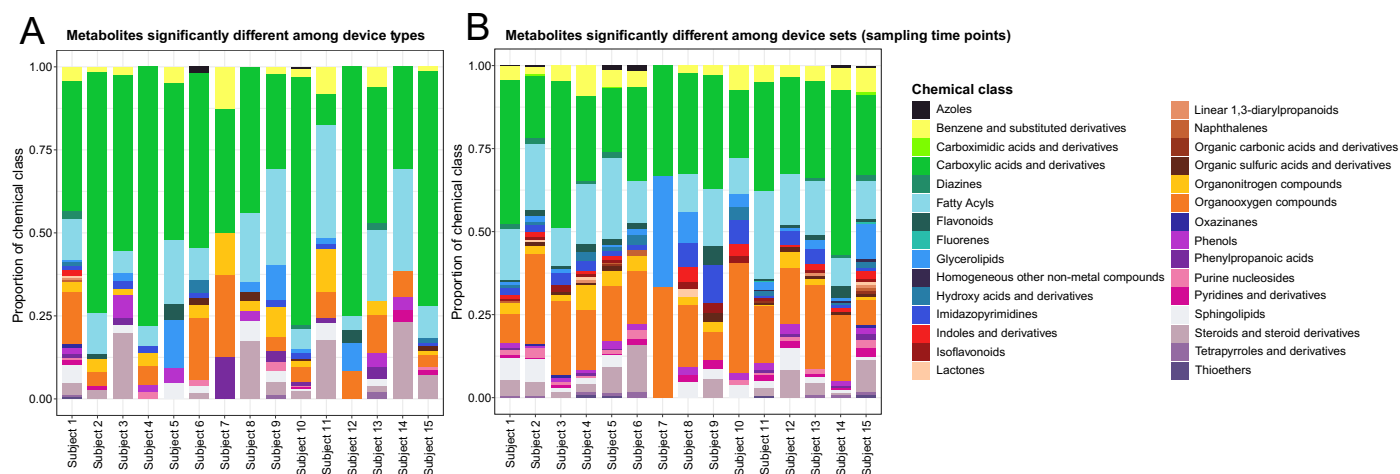
**Extended Data Fig. 5 | Chemical classes that differed significantly between the proximal and distal upper intestine.** Chemical enrichment statistics were performed by ChemRICH. Metabolites detected in >50% of intestinal samples were included in this analysis ( $n = 1182$  metabolites). These results were visualized by separating classes by chemical lipophilicity (logP) and chemical class significance level of  $-\log_{10}(p\text{-value})$ . Red circles indicate that the chemical

class was higher in the distal compared to proximal upper intestine, and blue indicates that the chemical class was lower in the distal compared to the proximal upper intestine. Purple indicates the chemical cluster has metabolites that are significantly higher as well as metabolites that are significantly lower in the distal compared to proximal upper intestine. Circle size represents the size of the chemical class.



**Extended Data Fig. 6 | Comparison of temporal and personal variation to spatial variation in the intestinal metabolome. a,** The number of metabolites that differed significantly different between intestinal regions (device types) or between meals (sampling time points of four capsule devices) calculated for each subject by analysis of variance (Kruskal-Wallis). Only metabolites detected in >50% of samples for each subject were used for this analysis (n = 1182 metabolites). Non-FDR-corrected  $p < 0.05$  was used as a significance threshold.

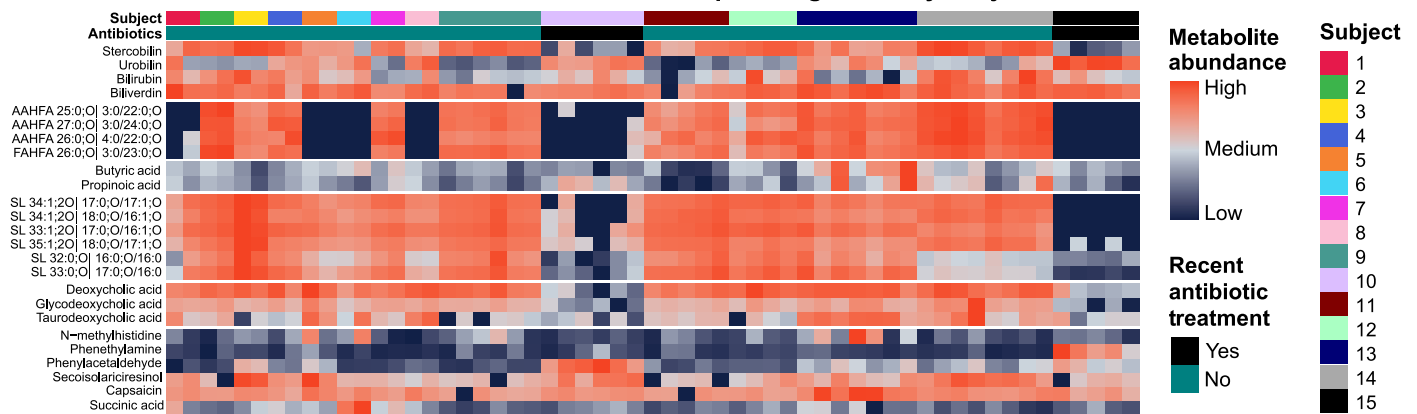
**b,** Multivariate discriminant analysis (PLS-DA) was performed to identify metabolites that were most important for distinguishing between subjects, or between regions. The 100 metabolites most important for distinguishing these groups were ranked by variable importance in projection score (VIP) and are categorized by chemical subclass. Chemical subclasses with <3 metabolites are reported as 'Other'.



**Extended Data Fig. 7 | The number of metabolites that were significantly different between intestinal regions (device types) or between meals (sampling time points of sets of 4 devices) were calculated for each subject by analysis of variance (Kruskal-Wallis). Only metabolites detected in >50% of samples for each subject were used for this analysis (n = 1182 metabolites). Non-FDR-corrected  $p < 0.05$  was used as a significance value cutoff.**

**a.** Metabolites with significantly different abundance between intestinal regions for each subject, grouped by chemical class and the proportion of each chemical class. **b.** Metabolites with significantly different abundance between sampling time points, grouped by chemical class and the proportion of each chemical class.

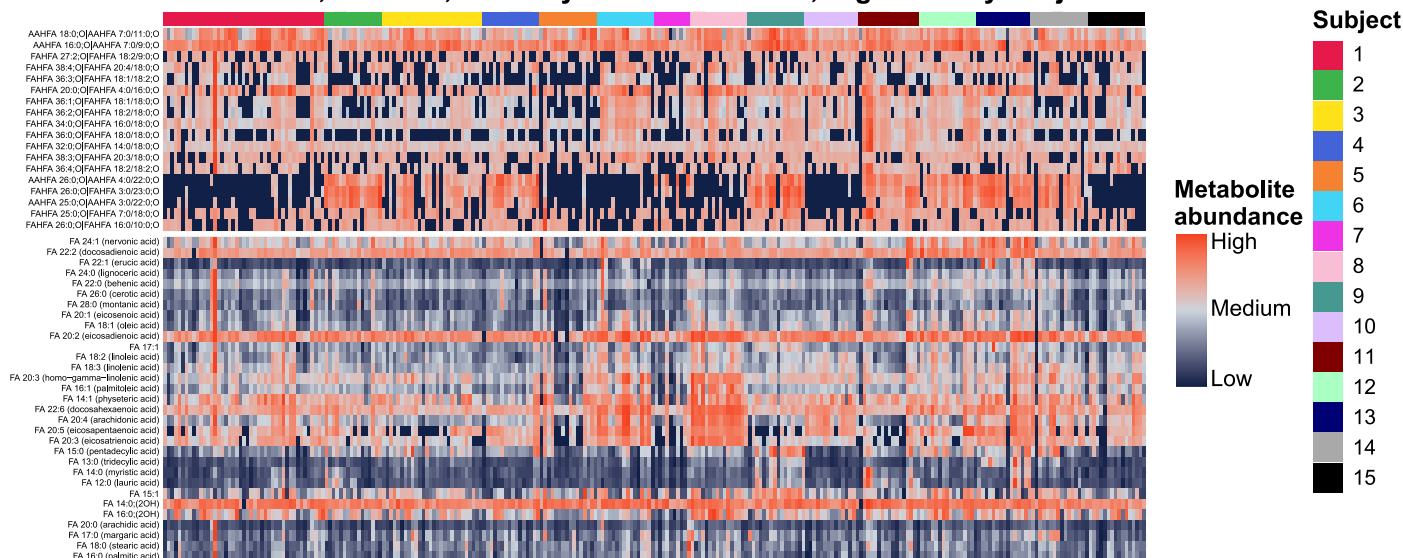
## Metabolite abundance in stool samples; organized by subject

**Extended Data Fig. 8 | 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. Data from all stool samples are shown, and samples are organized by

subject. The two subjects that consumed antibiotics 1 and 5 months prior to this study are highlighted. Color bar represents metabolite abundance (peak height) or concentration (ng/mL) for bile acids. Minimum and maximum values were used to set the color scale for each metabolite (each row).

## FAHFAs, AAHFAs, and fatty acid abundances; organized by subject



**Extended Data Fig. 9 | FAHFAs, AAHFAs, and fatty acids with strong inter-subject abundance differences.** Metabolites are fatty acid esters of hydroxy fatty acids (FAHFAs and AAHFAs) and fatty acids detected >50% of all device samples. All device samples are shown, and are organized by subject. Within the

top (FAHFA) and lower (fatty acid) sections, the metabolites are ordered based on hierarchical clustering. Color bar represents metabolite abundance (peak height) or concentration (ng/mL) for bile acids. Minimum and maximum values were used to set the color scale for each metabolite (each row).

**Extended Data Table 1 | The 20 chemical classes that exhibited the most significant differences between the proximal and distal upper intestine**

Cluster name	Cluster size	Cluster <i>p</i> -value	FDR <i>p</i> -value	Example compound	Metabolites increased from proximal to distal location	Metabolites decreased from proximal to distal location
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 phosphatidylethanolamines	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-04	2.2E-03	acetic acid	5	0

Significance was calculated by ChemRICH chemical enrichment statistics using univariate statistical results from an LMM as input (n = 1182 metabolites). Cluster size is the total number of metabolites within that chemical class.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

MS-DIAL v. 4.80 was used to process untargeted LC-MS/MS data. MS/MS spectra were matched to MassBank of North America, MS-DIAL and NIST20 spectral libraries. For GC-TOF MS data, Leco ChromaTOF v4.0 was used for data acquisition and primary data processing. Compounds were annotated by FiehnLib libraries used from MassBank.us. For targeted analyses of bile acids, Skyline software vs. 22.2. For targeted quantification of short chain fatty acids, Agilent MassHunter Quant v. B.09.00 was used. Retip software vs 1.0 was used for prediction of retention times. Genomic sequence data were de-multiplexed using the Illumina bcl2fastq algorithm. DADA2 software was used using pseudo-pooling with truncLenF and truncLenR parameters set to 250 and 180. Taxonomy was assigned using the Silva rRNA database v. 132.

Data analysis

Statistical tests were performed using R v. 4.0.3. Linear mixed effect models (LMM) were performed using the lmerTest and lme4 R packages. Benjamini-Hochberg was used to account for multiple-hypothesis testing. ChemRICH software vs. 1.0 was used to calculate enrichment statistics. Clustering was performed using the hclust function with the metabolite Spearman rank correlation matrix calculated using the cor function in R and Euclidean distance calculated with the as.dist function in R. Custom R scripts for metabolome data analyses are available from DOI 10.5281/zenodo.7659119. PLS-DA and PCA were performed with the ropls package in R. PLS-DA models to distinguish subject and device type had  $Q^2Y > 0.15$  and  $p < 0.05$  determined through 20 random permutations of class labels performed by the ropls R package.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.



## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Raw mass spectrometry data is available on the Metabolomics Workbench (<https://www.metabolomicsworkbench.org/>) under studies ST002073, ST002075, ST002407, ST002409 and ST002411. The 16S and metagenomics sequencing reads are available on NCBI under BioProject PRJNA822660. Taxonomy was assigned using the Silva rRNA database v. 132. Mass spectra were annotated using MassBank.us public libraries and NIST20 licensed libraries.

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

8 women and 7 men were enrolled (gender by self determination). Enrollment targeted roughly equal distribution between men and women, which was achieved.

Population characteristics

Age mean 42, range 22-64 years. BMI mean 23, range 19-31. Details on age and BMI are given in Supplement Table S3. Two participants had taken antibiotics within the past six months. Further exclusion and inclusion criteria given in the text.

Recruitment

George Triadafilopoulos at Silicon Valley Neurogastroenterology and Motility Center, Mountain View, CA 94040, USA, recruited subjects and obtained consents and samples. Participants were all healthy community-dwelling volunteers in the local area of the study site who had first or second-hand knowledge of the study and expressed their desire to participate. We did not need to advertise in order to recruit these 15 subjects, nor did we limit recruiting to a homogeneous subject population such as graduate students. While some self-selection bias may be present in that these participants were eager to participate in the study, all subjects met the objective inclusion and exclusion criteria set out beforehand. We did not bias study participation by including or excluding subjects based on their dietary preferences. Furthermore, the subjects represented a gender balance of 8 females and 7 males, as well as a good range of ages (mean 44 years old, range 22 to 67) and BMIs (mean 23, range 19 to 31).

Ethics oversight

The study was approved by the WIRB-Copernicus Group IRB (study #1186513) and informed consent was obtained from each subject. The full name and address of the IRB that reviewed and approved the protocol is WCG IRB 1019 39th Ave., SE, Suite 120 Puyallup, WA 98374.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

We chose our sample size based on the following rationale. Fifteen subjects swallowing up to 17 CapScan devices each provides a total of up to 255 device ingestions. If all subjects ingest the intended number of CapScan devices, the study has at least 80% statistical power to declare that more than 80% of the CapScan devices provided a viable liquid sample of intestinal luminal contents (that is  $H_0: \pi_{\text{CapScan}} \leq 0.80$  versus  $H_0: \pi_{\text{CapScan}} > 0.80$ ) using the exact binomial test, assuming that at least 86.7% of the recovered CapScan devices provide a viable sample. Setting aside the extra devices ingested by subject 1, out of the 255 CapScan devices ingested by the 15 subjects, 210 CapScan devices provided sufficient liquid sample to enable DNA sequencing (see companion Nature paper preprint doi:10.1101/2022.01.19.476920), representing a sampling success rate of 82.3%. Therefore, the sample size calculation performed a priori provided adequate power to detect the observed sampling success rate. Furthermore, with 255 device ingestions, the study has at least a 95% chance of observing at least one incident of a serious device related event when the underlying incidence of the event is at least 1.17% per ingestion. The most likely serious adverse event is device retention in the intestines. Capsule endoscopy devices are retained in the intestines at a rate of around 1.4% (Gastrointest Endosc 2010, 71:280286). Capsule endoscopes are approximately twice as large as the CapScan device. We therefore assumed that the chance of CapScan retention was at most 1% and that this 15-subject first-in-human pilot study had a 95% chance of detecting this safety endpoint.

Data exclusions

no data were excluded.

Replication	This is a first study on metabolome variance in human intestinal samples. Four samples were taken for each subject, between meals during two days of this study. Reproducibility of data acquisition and data processing was ensured by using 35 GI tract pooled QC samples. Data are given in Table S1, and visualized in Extended Figure 1A. On average across the three QC batches (Table S1), metabolites showed 24% relative standard deviation as technical error. Multivariate data variance analysis and density of cluster as given in Figure 1A shows that data reproducibility was excellent, with technical errors far smaller than the biological variance in the 255 capsules plus stool samples.
Randomization	Statistical treatments were not allocated to categories or study groups. Hence, no randomization towards any study characteristic (age, BMI, sex) was needed as the purpose of the study was to find out differences between temporal and spatial variance in metabolome data. Data acquisition on mass spectrometers was fully randomized across samples and across time points or sampling locations.
Blinding	Data were collected fully blinded to age, BMI, sex or other subject-relevant covariates such as antibiotics use. We did not use age, BMI, sex or other subject-relevant covariates. Hence, statistical analyses remained blinded for subject metadata.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging