scientific reports

OPEN

Longitudinal profiles of the fecal metabolome during the first 2 years of life

Elizabeth A. Holzhausen[1], Natalie Shen[2], Bridget Chalifour[1], ViLinh Tran[3], Zhenjiang Li[2], Jeremy A. Sarnat[2], Howard H. Chang[2], Dean P. Jones[3], Michael I. Goran[4], Donghai Liang[2] & Tanya L. Alderete[1]✉

During the first 2 years of life, the infant gut microbiome is rapidly developing, and gut bacteria may impact host health through the production of metabolites that can have systemic effects. Thus, the fecal metabolome represents a functional readout of gut bacteria. Despite the important role that fecal metabolites may play in infant health, the development of the infant fecal metabolome has not yet been thoroughly characterized using frequent, repeated sampling during the first 2 years of life. Here, we described the development of the fecal metabolome in a cohort of 101 Latino infants with data collected at 1-, 6-, 12-, 18-, and 24-months of age. We showed that the fecal metabolome is highly conserved across time and highly personalized, with metabolic profiles being largely driven by intra-individual variability. Finally, we also identified several novel metabolites and metabolic pathways that changed significantly with infant age, such as valerobetaine and amino acid metabolism, among others.

For infants, the first 1000 days represent a critical developmental window, during which metabolic, endocrine, neural, and immune systems are rapidly maturing. In tandem, the infant gut microbiome is also rapidly developing, until 2–3 years of age, when it reaches a more adult-like configuration[1]. This early colonization of the infant gut has far-reaching health implications, affecting immune system development[2], asthma and allergy[3–5], obesity and rapid growth[6–9], and cognitive development[10]. While the exact mechanisms by which the gut microbiome impacts human health are still being characterized, gut bacteria can impact host physiological systems through the production of metabolites that can have systemic effects[11–13]. High resolution metabolomics is an emerging, analytical, omics-based technology which offers global detection and characterization of the human metabolome, affording insights into how the metabolome interacts with both exogenous and endogenous exposures, and downstream health implications.

To date, several studies have examined the impact of mode of delivery, antibiotic usage, and early life feeding practices on the infant fecal metabolome. For example, 6-week old infants born vaginally compared to caesarian section (CS) were shown to have enriched metabolic pathways related to carbohydrate metabolism including glycolysis/gluconeogenesis and glyoxylate and dicarboxylate metabolism[14]. Amino acid metabolism has also been shown to be pronounced in infants exposed to antibiotics during the first weeks of life[16,17] as well as among formula-fed infants, compared to breastfed infants[14]. Further, breastfed infants have different metabolic profiles compared to their formula-fed counterparts at 3-, 6-, and 9-months of age, including differences in butyric acid, d-sphingosine, betaine, and kynurenic acid[18]. However, few studies have characterized the fecal metabolome using frequent repeated sampling during the first 2 years of life. This information is critical to understanding immune system development as infants increase food diversity and gain mobility.

While prior research offers important insight into the development of the fecal metabolome, previous longitudinal studies followed infants only through the first year of life[15–20], and none have been conducted in a Latino cohort. The aim of this study was to contribute to the existing literature by systematically describing the development of the infant fecal metabolome during the first 2 years of life. Specifically, we examined the fecal metabolome at 1-, 6-, 12-, 18-, and 24-months of age in 101 Latino infants from the Southern California Mother's Milk Study – a well-established cohort that was recruited based on an intention to breastfeed for at least three months. As a secondary aim, we sought to assess which factors were important predictors of overall fecal metabolites and fecal metabolomic pathways, including mode of delivery, antibiotic exposure, and infant age. Future work in this

[1]Department of Integrative Physiology, University of Colorado – Boulder, Boulder, CO, USA. [2]Rollins School of Public Health, Emory University, Atlanta, GA, USA. [3]School of Medicine, Emory University, Atlanta, GA, USA. [4]Children's Hospital Los Angeles, Los Angeles, CA, USA. ✉email: Tanya.Alderete@colorado.edu

cohort will expand on this initial analysis to examine how breastfeeding, formula feeding, solid food introduction, and environmental exposures impact fecal metabolome throughout the first 2 years of life.

Results

Study population characteristics. General population characteristics are shown in Table 1. At the 1-month postpartum visit, mothers were 29 ± 6 years old (18–45), and most had a body mass index (BMI) in the overweight (33%) or obese (40%) category. The average infant age at fecal metabolome assessment was 1.1, 6.2, 12.3, 18.3, and 25.1 months, respectively, at planned 1-, 6-, 12-, 18-, and 24-month collection[21]. Half of the infants were female, 74% were born vaginally, and 10.9% received antibiotics in the first 2 years of life. Most (97%) were breastfed at 1-month, and 41% were breastfed at 24-months of age.

Temporal trends in fecal metabolites. Overall, there were 11,345 metabolic features extracted from the HILIC and 8,609 in the C18 chromatography after removal of features that were present in less than 10% of samples. From this, we confirmed the chemical identities of 143 unique metabolites from HILIC chromatography and 104 metabolites from C18 chromatography with Level 1 evidence (i.e., features whose *m/z* and retention time could be matched to authentic standards with MS/MS under identical conditions). Among the confirmed metabolites, many were highly conserved over time. For instance, 115/143 metabolites in the HILIC column were observed in at least 50% of samples at each timepoint, as were 78/104 metabolites in the C18 column (Fig. 1). As a sensitivity analysis, we additionally explored metabolites that were present in 25% and 75% of samples (Supplemental Fig. 1 and Supplemental Fig. 2). We found that there were 135 and 97 confirmed metabolites that were present in at least 25% of samples at each timepoint in the HILIC and C18 column, respectively. There were 72 and 50 metabolites present in 75% or more of samples at every timepoint in the HILIC and C18 columns, respectively.

Observed metabolites included (Figs. 2 and 3) amino acids (e.g., tyrosine, tryptophan, methionine, isoleucine, aspartate, proline), amino acid metabolites (e.g., phenylacetate, oxoproline, taurine, cystathionine), purines (adenine, guanine, hypoxanthine), pyrimidine (thymine) and vitamin metabolites [thiamine (B1), nicotinamide (B3), pyridoxine (B6), pyridoxate, alpha-tocopherol (E), dethiobiotin]. Lipid metabolites included sphingosine, sphinganine, choline, carnitine, cholesterol, free fatty acids (e.g., linoleate, oleate) and lysophosphatidyl choline. Microbial metabolites with known adverse activities (indoxyl sulphate, valerobetaine) were present, and dietary (cinnamaldehyde) and environmental chemicals (pirimicarb) were also detected.

Changes in fecal metabolites were largely driven by age and intra-infant variability. Given that the intensities of many confirmed metabolites changed during the first 24-months of life (Fig. 2 and 3), we performed non-parametric univariate permutational multivariate analysis of variance tests (PERMANOVA) to assess which variables explained the most variance in overall metabolite composition. We found that metabolic variation was largely driven by intra-individual variability such that 29.2% ($P = 0.001$) of the variability in confirmed metabolites in the HILIC column, and 30.0% ($P = 0.001$) of the variability in confirmed metabolites in the C18 column could be attributed to the infant providing the sample.

In addition to individual variability, we also examined the impact of infant sex, antibiotic exposure (i.e., having received any antibiotics since birth), and infant age in days on overall profiles of fecal metabolites.

	Mean $\pm$ SD or N, % N = 101
Maternal characteristics	
Age (years) at 1-month postpartum visit	29 ± 6
Socioeconomic status (SES)[a]	27 ± 12
Pre-pregnancy BMI (kg/m^2)	28.7 ± 6.0
Infant characteristics	
Age (days) at 1-month postpartum visit	32.8 ± 3.1
Sex (Female, Male, %Female)	51, 50, 50.5%
Age of solid foods (months)[b]	5.9 ± 1.7
Antibiotics (Yes, No, %Yes)	11, 90, 10.9%
Birth mode (Vaginal, C-section, %Vaginal)	75, 26, 74.3%
Birth weight (kg)	3.4 ± 0.4
Birth length (cm)	50.4 ± 2.5
Gestational age	
Early (< 38 weeks gestation)	26 (25.7%)
On time (38–42 weeks gestation)	54 (53.5%)
Late (> 42 weeks gestation)	21 (20.8%)

Table 1. Characteristics of 101 mother-infant dyads from the Southern California Mother's Milk Study, 2016–2019. Data reported are mean and standard deviation (SD) unless otherwise noted. [a] Median replacement was performed for missing values (N = 2). [b] N = 100. SES, socioeconomic status; BMI, body mass index.
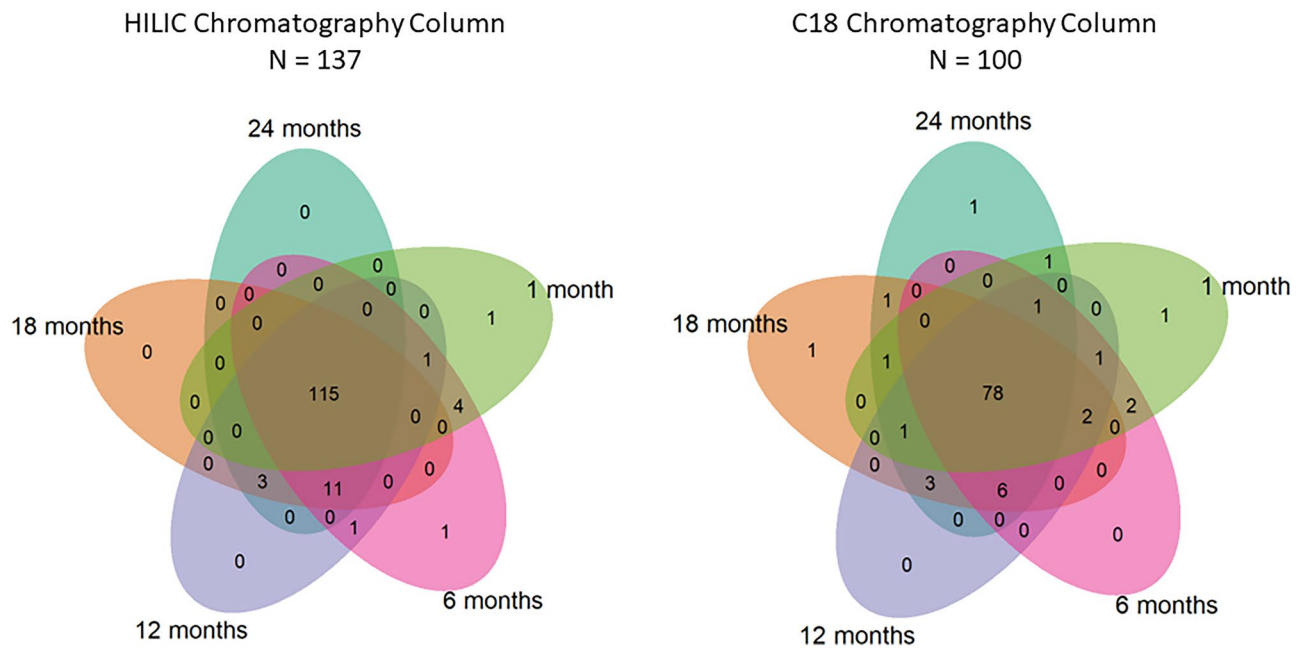
Figure 1. Confirmed metabolites observed longitudinally in at least 50% of samples during the first 2 years of life in the HILIC (left) and C18 (right) chromatography columns.

Infant sex and antibiotic exposure were not important predictors of metabolites in the HILIC column, with sex explaining just 0.3% of variability ($P = 0.1$) and antibiotic exposure explaining 0.4% of variability ($P = 0.02$). Similarly, infant sex explained 0.2% of variability in the metabolites in the C18 column ($P = 0.3$) and antibiotic exposure explained 0.3% of variability ($P = 0.1$). Despite this, infant age explained much more of the variability in fecal metabolites where 6.7% ($P = 0.001$) and 6.2% ($P = 0.001$) of the variability in confirmed metabolites in the HILIC and C18 columns were explained by infant age (Fig. 4). Indeed, after adjustment for multiple testing using the Benjamini-Hochberg (BH) procedure, we found there were 87 confirmed metabolites in the HILIC chromatography column and 77 in the C18 chromatography column which were significantly associated with infant age in days ($P_{BH} < 0.05$). Table 2 summarizes the top 10 most significant results for each chromatography column. These showed increases in some diet-derived metabolites (methyl vanillate, hydrocinnamic acid) and decreases in free fatty acids (arachidonate, laurate, homolinoleic acid). A complete list of statistically significant results can be found in Supplemental Table 2.

As shown in Fig. 5, we examined selected metabolites that were most significantly associated with infant age in days in the HILIC and C18 chromatography columns. Mean intensity of valerobetaine, a microbiome-derived metabolite associated with increased adiposity[22], increased by 588% between 1- and 24-months of age. Other confirmed metabolites in the HILIC column that decreased with infant age included kynurenine, a metabolite associated with immune cell function, whose mean intensity decreased by 79.1%. In the C18 column, mean intensity of the polyunsaturated fatty acids arachidonic acid and eicosadienoic acid decreased with infant age (an 88.5% and 85.5% decrease, respectively). In addition to examining the level of intensity over time, we also visualized the prevalence of each metabolite that was most significantly associated with infant age (Supplemental Fig. 3). While most of these metabolites were detected with high prevalence across all timepoints, valerobetaine prevalence increased dramatically over time and prevalence of arachidonic and docosahexaenoic acid decreased precipitously between 1- and 24-months.

Amino acid, bile acid, and carbohydrate metabolism pathways were enriched with infant age. Results from the linear mixed effects models were used to perform pathway enrichment analyses using Metapone[23]. There were 18 metabolic pathways significantly associated with infant age ($P \le 0.05$) and with weighted number of significant metabolites ≥ 1.5. Pathways associated with infant age are summarized in Fig. 6. Of the pathways associated with infant age, 7 pathways were related to amino acid metabolism, including both non-essential and essential amino acid pathways and the urea cycle. These pathways are linked to many of the metabolites in Table 2 (e.g., pyridoxine is the vitamin supporting amino acid nitrogen elimination through the urea cycle). Two pathways containing methionine are connected to individual metabolites described above (e.g., taurine), and to the bile acid pathway and the spermidine/spermine pathway. Seven pathways were related to carbohydrate metabolism, and one was related to bile acid metabolism. Other significant pathways included metabolism of xenobiotics and membrane transport (ATP-binding cassette transporters).

Figure 2. Longitudinal patterns of metabolite intensity of confirmed fecal metabolites detected in the HILIC chromatography column during the first 2 years of life. Intensity presented is standardized by dividing by intensity standard deviation, and ranges from 3.1 to 32.4, with dark purple representing the highest intensity and white representing the lowest intensity. Grey indicates a missing value. Metabolites detected in at least 80% of samples were included. Observations are grouped by visit, with the age in days corresponding with each visit as follows: 01 m (25–46 days), 06 m (164–219 days), 12 m (351–429 days), 18 m (517–582 days), and 24 m (709–916 days).

Discussion

This study provides important new information concerning the longitudinal development of the fecal metabolome in Latino infants from 1 to 24 months, a time period with far-reaching health implications for immune

Figure 3. Longitudinal patterns of metabolite intensity of confirmed fecal metabolites detected in the C18 chromatography column during the first 2 years of life. Intensity presented is standardized by dividing by intensity standard deviation, and ranges from 4.7 to 37.6, with dark purple representing the highest intensity and white representing the lowest intensity. Grey indicates a missing value. Metabolites detected in at least 80% of samples were included. Observations are grouped by visit, with the age in days corresponding with each visit as follows: 01 m (25–46 days), 06 m (164–219 days), 12 m (351–429 days), 18 m (517–582 days), and 24 m (709–916 days).

system development[2], asthma and allergy[3–5], obesity and growth[6–9], and cognitive development[10]. At least 50% prevalence at each timepoint was observed for about 200 confirmed metabolites, indicating that the fecal metabolome was relatively stable. Variability was primarily explained by the individual providing the sample as well
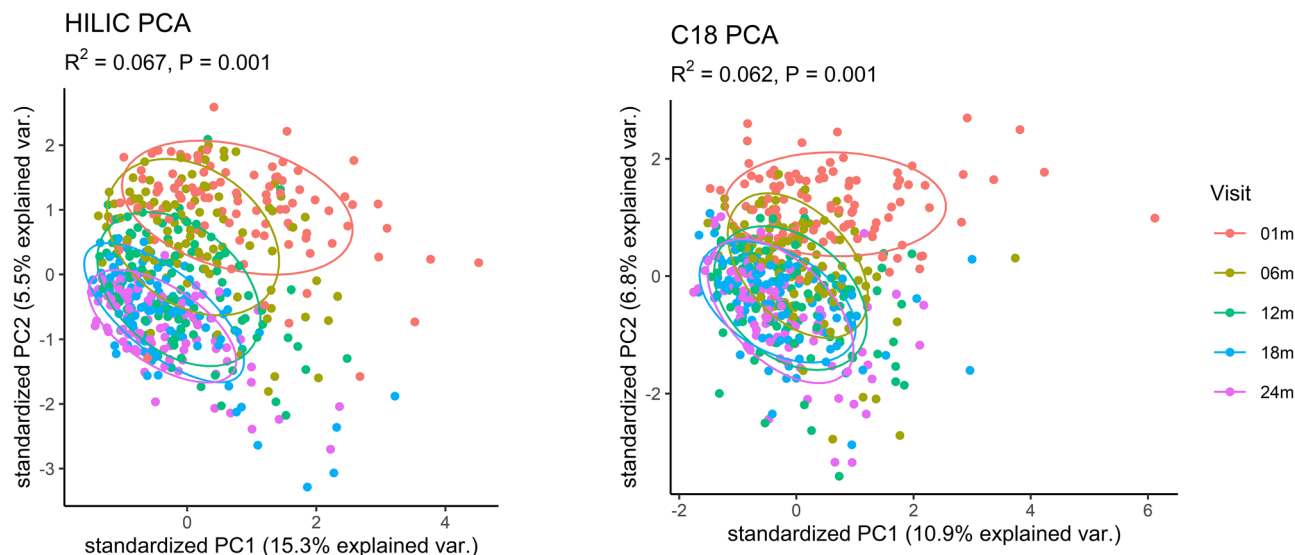
Figure 4. Ordination plot of principal components 1 and 2 for confirmed metabolites detected in HILIC (left) and C18 (right) by infant age in days. R^2- and P- values calculated using Permutational Multivariate Analysis of Variance, with age in days as the explanatory variable. Points are colored by visit, with the age in days corresponding with each visit as follows: 01 m (25–46 days), 06 m (164–219 days), 12 m (351–429 days), 18 m (517–582 days), and 24 m (709–916 days).

HILIC chromatography column			C18 chromatography column		
Metabolite	**Direction**	**P_{BH}**	**Metabolite**	**Direction**	**P_{BH}**
4-Pyridoxate	↑	1.2×10^{-47}	Arachidonic Acid	↓	2.3×10^{-34}
Dihydroxyacetophenone/ Resorscinol monoacetate/ Methylparaben	↑	5.7×10^{-55}	Docosahexaenoic acid	↓	1.8×10^{-34}
Hexadecanol	↑	4.2×10^{-18}	Eicosadienoic acid	↓	3.5×10^{-20}
Kynurenine	↓	1.3×10^{-26}	Gondoic acid	↓	3.1×10^{-17}
Methylnicotinium*	↑	8.3×10^{-27}	Homolinoleic acid	↓	2.3×10^{-18}
Methylhippurate	↑	1.0×10^{-17}	Hydrocinnamic acid	↑	1.3×10^{-18}
Methyl vanillate	↑	2.5×10^{-21}	Hypoxanthine	↓	1.0×10^{-15}
N-Acetyl-phenylalanine	↑	3.3×10^{-19}	Laurate	↓	1.6×10^{-16}
Valerobetaine	↑	2.7×10^{-35}	Mandelic acid/2-,3-Hydroxyphenylacetic acid	↑	3.0×10^{-41}
Xanthurenic acid	↑	3.6×10^{-22}	*N*-alpha-acetyl-L-lysine	↑	3.0×10^{-27}

Table 2. Top ten confirmed metabolites detected by HILIC and C18 chromatography columns that were most significantly associated with infant age, based on the results of linear mixed effects models with random intercepts for individual to account for repeated measures. Results from these models were adjusted using the Benjamini–Hochberg (BH) procedure. Overall, there were 87 metabolites significantly associated with infant age in days after adjustment for multiple testing in the HILIC chromatography column and 77 in the C18 chromatography column. *This compound was unidentified but is an accurate mass match for methylnicotinium.

as infant age. Metabolites associated with infant age included amino acid metabolites, such as kynurenine, and polyunsaturated fatty acids, arachidonic and docosahexaenoic acid. Of potential importance relative to weight gain and obesity, valerobetaine increased substantially over the first 24 months. Valerobetaine is a metabolite produced by the intestinal microbiome which was recently found to inhibit fatty acid oxidation by mitochondria[22]. In mice, valerobetaine causes increased adiposity and in humans, valerobetaine is associated with high BMI and fatty liver[22]. The latter association may be especially relevant to Latino health because adolescent Latino children have increased risk of non-alcoholic fatty liver disease (NAFLD)[24,25]. Other metabolites whose intensity was significantly associated with infant age were also associated with pathways related to amino acid metabolism, carbohydrate metabolism, bile acid metabolism, and lipid metabolism.

A previous study on preterm infants found that intra-individual variability explained 43% of variation in fecal metabolites during the first months of life; whereas, health outcomes (sepsis and necrotizing enterocolitis), delivery mode, and antibiotics were not significantly associated with metabolomic variation[26]. Several of the
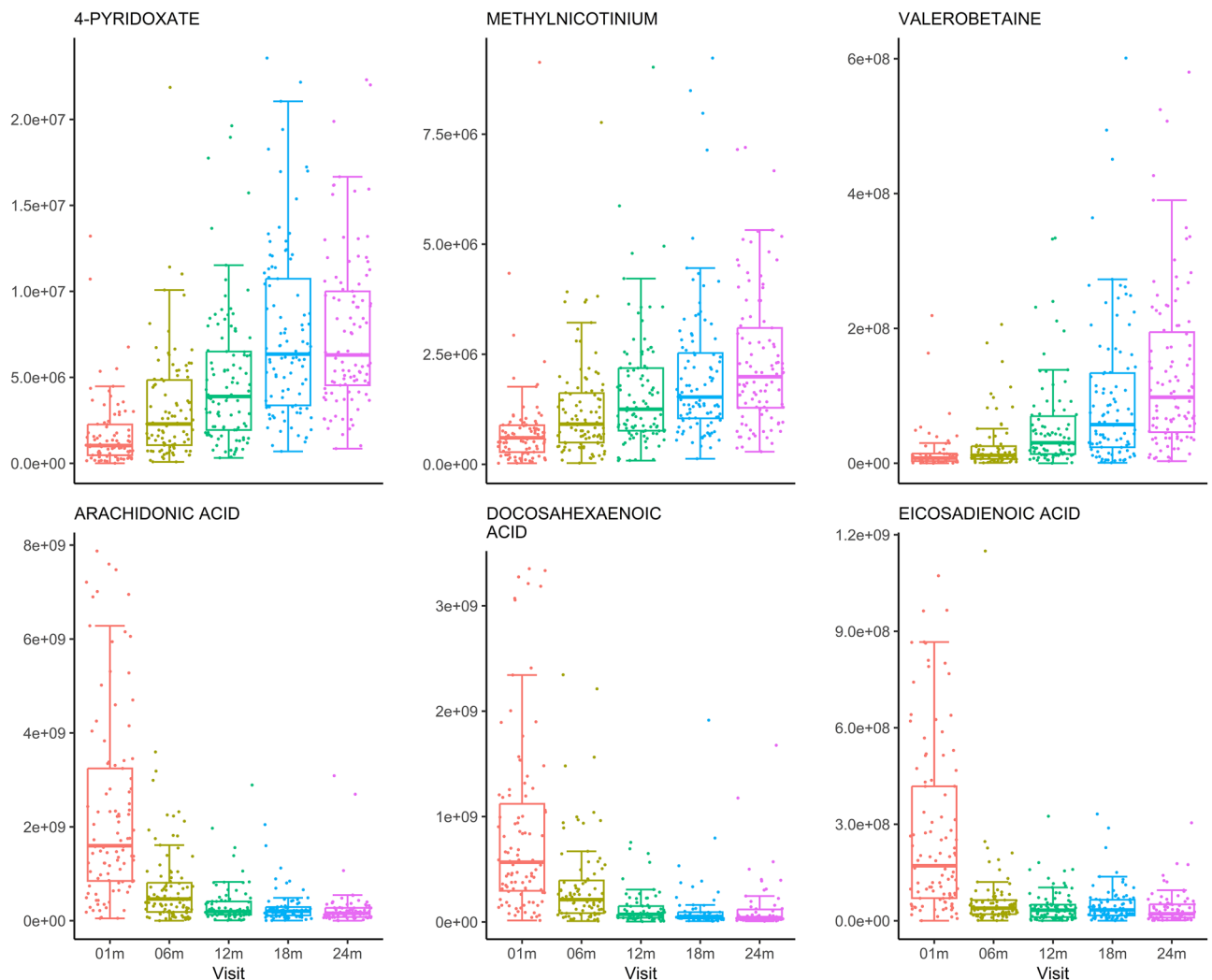
Figure 5. Visual representation of the intensity of selected confirmed metabolites detected in the HILIC (top) and C18 (bottom) chromatography columns which varied most significantly with infant age. Observations are grouped by visit, with the age in days corresponding with each visit as follows: 01 m (25–46), 06 m (164–219), 12 m (351–429), 18 m (517–582), and 24 m (709–916). The metabolite labeled methylnicotinium was unidentified but was an accurate mass match for methylnicotinium.

metabolites that we found to be related to infant age have also been identified previously. For example, we found that uracil, aspartate, methionine, tyrosine, phenylalanine, and valine were significantly increased with infant age, while glucose and fucose metabolites were significantly decreased. Interestingly, previous work has found these same metabolites to be associated with either formula or breast feeding[14]. We also observed that methyl vanillate, which is a compound identified in cow's milk[27] and also produced by Brewer's yeast (Saccharomyces cerevisiae)[28], significantly increased with infant age. Given this, the changes that we observed in metabolite intensity over time may reflect increased formula or cow's milk consumption, increased consumption of solid foods, and/or decreased breastfeeding or weaning as infants age. Several metabolites that changed significantly over time in our study are found in breast milk and infant formula, including kynurenine, which is present in both breast milk and formula, but is found in higher amounts in formula compared to breast milk[29]. Arachidonic acid and docosahexaenoic acid (DHA) are both long-chain polyunsaturated fatty acids that are found in breast milk[30]. Infant formulas are also supplemented with DHA[31]. These observations further bolster the hypothesis that shifts in the early-life fecal metabolome are driven largely by dietary alterations during the first 2 years of life. Given these observations, future work in this cohort will explore the associations between the fecal metabolome with early life feeding practices, the infant gut microbiota, and adiposity.

In this study, we found that pathways related to amino acid metabolism, carbohydrate metabolism, and bile acid metabolism were associated with increased infant age. Interestingly, several of these pathways have also been linked with infant breast or formula feeding[14]. For example, we observed that seven amino acid pathways increased with infant age, one of which (arginine and proline metabolism) has been shown to be enriched in the fecal metabolome of formula-fed infants compared to breastfed infants at 6-weeks of age[14]. We also found seven
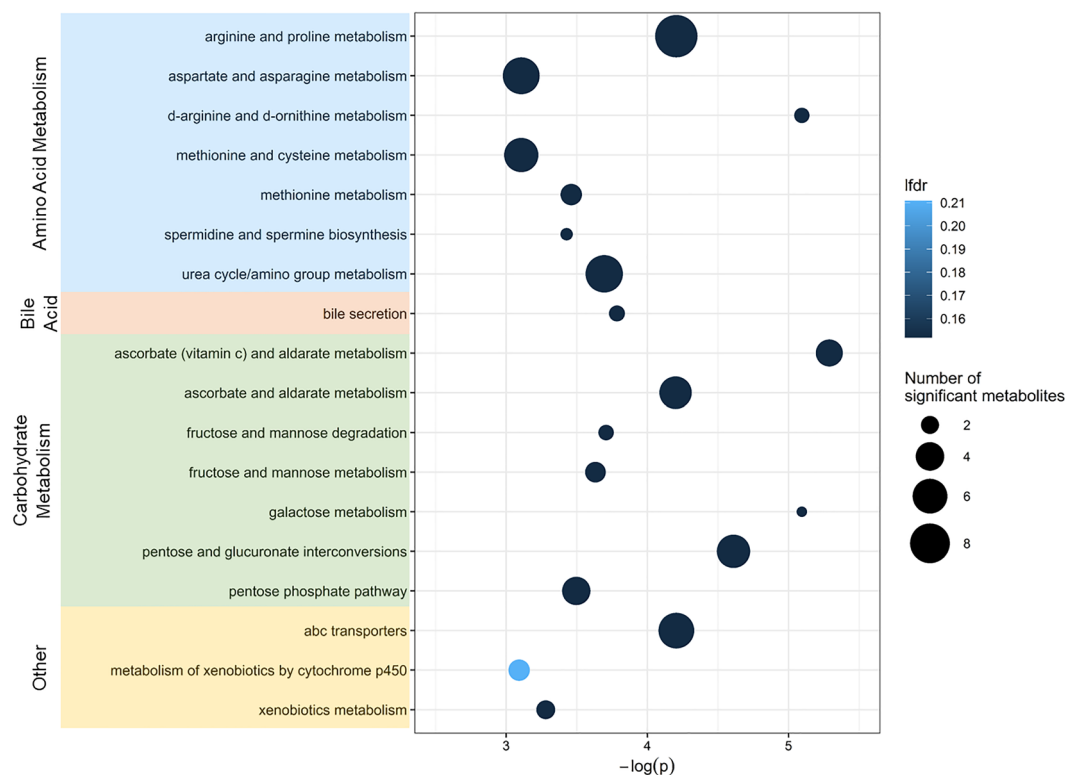
Figure 6. Overview of enriched pathways by confirmed HILIC and C18 chromatography column metabolites that were significantly associated with infant age, identified using Metapone. Pathways with P ≥ 0.05 and weighted number of significant metabolites ≥ 1.5 are included. Point size indicates the number of significant metabolites identified within each pathway and point color indicates significance after adjustment for multiple testing via conditional local false discovery rate (lfdr). Pathways are grouped by super pathway (amino acid metabolism, bile acid, carbohydrate metabolism, lipid, and other), indicated by color.

carbohydrate metabolism pathways enriched with infant age, two of which (galactose metabolism and fructose and mannose metabolism) were enriched in breastfed infants, compared to formula-fed infants[14].

Pathways that we identified in the current analysis have also been linked with antibiotic exposure and mode of delivery in other fecal metabolomics studies that investigated infants. For example, amino acid metabolism has previously been shown to be enriched in infants exposed to antibiotics during the first weeks of life[16,17]. We also found metabolites linked to ABC transporters were enriched with increasing infant age. Previous studies have observed fecal metabolome enrichment of this pathway among infants aged approximately 15-weeks and born cesarean-section, compared to vaginally-born infants[15] and among preterm infants with antibiotic exposure during the first 14 days of life, compared to preterm infants who did not receive antibiotics[16]. Lastly, one carbohydrate metabolism pathway (galactose metabolism), which we found was enriched with infant age, has previously been shown to be enriched in 6-week old infants who were born by cesarean section[14].

While this study had several strengths, including repeated sampling and comprehensive metabolic profiling of the infant fecal metabolome in a well-established cohort of infants, some limitations are worth noting. First, stool was collected using OMNIGene GUT kits, which may have decreased the number of identified metabolites[33]. However, previous studies have found that a much higher proportion of variability in the fecal metabolome is attributable to individual, compared with collection method[32]. Given the young age of our participants, we were unable to collect blood samples, which limited our ability to examine the circulating metabolome. Therefore, future studies should incorporate fecal metagenomics and serum metabolomics since gut bacteria likely play an important role in shaping the fecal metabolome. Additionally, as this study was focused on metabolites that were identified with Level 1 evidence, we largely characterized patterns in a relatively limited number of endogenous metabolites over the first 2 years of life. This study was also conducted in an exclusively Latino cohort with exclusions such as preterm birth or low birth weight, cigarette smoking or recreational drug use, which may limit the generalizability of these findings to the broader population. In particular, given that our cohort was comprised of Latinos we were unable to assess whether race/ethnicity is an important predictor of the fecal metabolome. Future work is needed to replicate these findings in other populations. Nevertheless, we detected numerous metabolites that were previously observed in other infant populations. Lastly, exploring feeding practices in this cohort is complex because mothers were recruited based on an intention to breastfeed. Furthermore, while weaning may impact the fecal metabolome, we were unable to address weaning in this study due to incomplete information. Therefore, while examining infant feeding was beyond the scope of the current study, future work will be needed

to tease apart the relative intake of formula and breast milk over time and characterize how breastfeeding, formula feeding, and solid food introduction impact fecal metabolome throughout the first 2 years of life.

Conclusions

This study characterizes the development of the infant fecal metabolome during the first 2 years of life. Overall, we found evidence that the metabolome was relatively stable across time and highly personalized, with metabolic profiles being largely driven by intra-individual variability. We also identified several novel metabolites, such as valerobetaine, and pathways, including amino acid biosynthesis, which were significantly associated with infant age.

Methods

Study population.   The Southern California Mother's Milk Study is an ongoing, longitudinal cohort of 219 Latino mother-infant pairs who were recruited between 2016 and 2019 from Los Angeles County maternity clinics, which has been previously described[21]. Participants were eligible to participate if they were ≥ 18 years old at time of delivery; had a healthy, singleton birth; enrolled in the study by 1-month postpartum; and could read at a 5th grade level in either Spanish or English. Potential participants were excluded if they had any diagnoses known to impact mental/physical health, nutritional status, or metabolism; were currently using tobacco or recreational drugs; had infants who were pre-term or low birth weight; or had infants with clinically diagnosed fetal abnormalities. The Institutional Review Boards of the University of Southern California, Children's Hospital Los Angeles, and the University of Colorado Boulder approved of the study procedures and all research was performed in accordance with the relevant guidelines and regulations. Written informed consent was obtained from participants at time of enrollment.

Study design.   Participants were enrolled by 1-month postpartum and attended follow-up visits at 6-, 12-, 18-, and 24-months postpartum. 219 mother-infant dyads were initially enrolled in the Mother's Milk cohort. As previously reported[34], the primary aim of the Mother's Milk Study was to assess the impact of sugars and human milk oligosaccharides on the infant microbiome and obesity. Briefly, socioeconomic status was estimated using a modified Hollingshead index, as previously described[21,35]. Maternal self-report was used to classify infants' gestational age as early (< 38 weeks gestation), on time (38–42 weeks gestation), and late (> 42 weeks gestation). Questionnaires were used to determine birth mode, infant antibiotic exposure, and infant feeding practices. Additional funding supported the analysis of 600 fecal metabolomics samples in this cohort and a subset of 127 participants were selected to undergo fecal metabolomics analysis, to maximize the number of participants with repeated fecal metabolome samples. 101 infants had complete fecal metabolomics samples at all 5 visits and were included in this analysis. Those individuals that were excluded due to missing data did not differ significantly from those who were included in the analysis (Supplemental Table 1).

High-resolution metabolomics.   OmniGene GUT kits were used to collect infant stool samples at 1-, 6-, 12-, 18-, and 24- months of age. Untargeted high-resolution metabolomics analysis was carried out by the Emory Clinical Biomarkers Laboratory, as previously described[36,37]. To precipitate proteins, stool samples were first added to ice-cold acetonitrile. Samples were then kept on ice for 30 min, centrifuged for 10 min at 14,000 g, and kept at 4 °C until analysis. Extractants were examined in triplicate using high-resolution mass spectrometry (LC-HRMS) (Dionex Ultimate 3000, Thermo Scientific Orbitrap Fusion).

Instrumentation and analytical conditions.   Hydrophilic interaction liquid chromatography (HILIC) (Waters XBridge BEH Amide XP HILIC column; 2.1×50 mm^2, 2.6 μm particle size) with positive electrospray ionization (ESI) and reverse phase (C18) chromatography (Higgins Targa C18 2.1×50 mm^2, 3 μm particle size) with negative ESI were used. HILIC analyte separation was conducted using water, acetonitrile, and 2% formic acid mobile phases following the subsequent gradient elution. The initial 1.5-min period consisted of 22.5% water, 75% acetonitrile, and 2.5% formic acid followed by a linear increase to 75% water, 22.5% acetonitrile, and 2.5% formic acid at 4 min, followed by a final hold for 1 min. Analyte separation for the C18 chromatography column was conducted using water, acetonitrile, and 10 mM ammonium acetate mobile phases under the following gradient elution. The initial 1-min period consisted of 60% water, 35% acetonitrile, and 5% ammonium acetate followed by a linear increase to 0% water, 95% acetonitrile, and 5% ammonium acetate at 3 min with a final hold for the last 2 min. Mobile phase flow rate was 0.35 mL/min for the first minute and was increased to 0.4 mL/min for the last 4 min for both the HILIC and C18 chromatography columns. LC-HRMS was run in full scan mode, with 120 k resolution and had a range of mass-to-charge ratio (*m/z*) from 85 to 1,275. Tuning parameters for sheath gas were 45 (arbitrary units) for positive ESI and 30 for negative ESI. For positive ESI, auxiliary gas was set to 25 (arbitrary units) and spray voltage was set at 3.5 kV, and for negative ESI, auxiliary gas was set to 5 and spray voltage was set to − 3.0 kV. Internal standards included pooled stool and standard reference materials for human metabolites in stool. These internal standards were added at the beginning and end of each 20-sample batch for quality control and standardization.

Metabolite confidence and identification.   Data from positive and negative ion modes were analyzed separately, and raw files were converted to the .mzXML format. Then, metabolomic signals (i.e., metabolic features) were extracted and aligned using apLCMS with modification of xMSanalyzer for quality control and reduction of batch effects following instrument analysis[38,39]. The coefficients of variation (CV) of metabolites were assessed as part of quality control. Metabolites whose intensity had CV > 30% were removed and intensi-

ties of metabolic features were averaged across triplicates. Metabolic features which were detected in < 10% of samples were excluded. Outliers were assessed visually using principal component analysis (PCA) of the $\log_2$ transformed metabolite intensities. In a sensitivity analysis, samples with PCA score was > 3 standard deviations for PC 1 or PC 2 were removed (not shown). However, there were no important differences in results, so these observations were not removed. Metabolomic features were then annotated and confirmed using the Metabolomics Standards initiative criteria[40]. Level 1 confidence was assigned to features whose *m/z* and retention time matched the authentic standards analyzed with MS/MS under identical conditions (within 10 ppm and 50 s).

Statistical analysis. Descriptive statistics for key variables were performed on the full analytic data set. We used the VennDiagram package in R to visualize how many metabolites were present in 25%, 50%, and 75% of samples at each visit[41]. Heatmaps to visualize $\log_2$ transformed metabolite intensities over time were generated using the ComplexHeatmap package in R[42]. Next, we performed PCA on the $\log_2$ transformed metabolite intensities to visualize overall metabolite profiles over time. We used permutational multivariate ANOVA (PERMANOVA) tests to explore how overall fecal metabolite intensities changed over time, and in relation to individual, infant sex, and antibiotic exposure using the adonis2 function implemented by the vegan package in R, using Euclidian distance (permutations = 1000)[43]. Linear mixed effects models were then used to estimate the relationship between the $\log_2$ transformed intensity of each confirmed metabolite and infant age in days using the lme4 package in R[44]. Models included random intercepts to account for repeated measures and were adjusted for multiple testing using the Benjamini–Hochberg procedure[45]. Boxplots were used to visualize the intensity of 6 selected confirmed metabolites associated with infant age in days in the HILIC and C18 chromatography columns by visit. Finally, we used the Metapone package in R[23] to jointly assess which metabolic pathways were associated with infant age in days, with combined linear mixed effects model results generated for the HILIC and C18 chromatography columns, using a weighted gene set enrichment analysis (GSEA) modified to untargeted metabolomics data. Briefly, Metapone employs established online MS databases of metabolites to putatively annotate metabolic features, incorporates a weight to limit the influence of single features that are annotated to multiple metabolites, and finally applies the modified GSEA test to focus on groups of metabolic features playing a role in the same biological pathways[23]. Metapone uses local false discovery rate (lfdr) to adjust for multiple testing, which is a Bayesian approach with minimal a priori assumptions and which does not rely on the assumption that individual statistical tests are independent[46].

Data availability
Data cannot be shared publicly because they include potentially identifying information on human subjects. The data that support the findings of this study are available upon reasonable request from the corresponding author, TLA.

References
1. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**(7402), 222–227 (2012).
2. Francino, M. P. Early development of the gut microbiota and immune health. *Pathogens.* **3**(3), 769–790 (2014).
3. Arrieta, M. C. *et al.* Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci. Trans. Med.* **7**(307), 307ra152 (2015).
4. Stiemsma, L. T. *et al.* Shifts in *Lachnospira* and *Clostridium* sp. in the 3-month stool microbiome are associated with preschool age asthma. *Clin. Sci.* **130**(23), 2199–207 (2016).
5. Azad, M. B. *et al.* Infant gut microbiota and food sensitization: associations in the first year of life. *Clin. Exp. Allergy* **45**(3), 632–643 (2015).
6. Korpela, K. *et al.* Childhood BMI in relation to microbiota in infancy and lifetime antibiotic use. *Microbiome.* **5**(1), 26 (2017).
7. Kalliomäki, M., Carmen Collado, M., Salminen, S. & Isolauri, E. Early differences in fecal microbiota composition in children may predict overweight. *Am. J. Clin. Nutr.* **87**(3), 534–538 (2008).
8. Vael, C., Verhulst, S. L., Nelen, V., Goossens, H. & Desager, K. N. Intestinal microflora and body mass index during the first 3 years of life: an observational study. *Gut Pathogens.* **3**(1), 8 (2011).
9. Alderete, T. L. *et al.* Early life gut microbiota is associated with rapid infant growth in Hispanics from Southern California. *Gut Microbes.* **13**(1), 1961203 (2021).
10. Carlson, A. L. *et al.* Infant gut microbiome associated with cognitive development. *Biol. Psychiat.* **83**(2), 148–159 (2018).
11. Ma, N. & Ma, X. Dietary amino acids and the gut-microbiome-immune axis: physiological metabolism and therapeutic prospects. *Comp. Rev. Food Sci. Food Saf.* **18**(1), 221–242 (2019).
12. Kriaa, A. *et al.* Microbial impact on cholesterol and bile acid metabolism: current status and future prospects. *J. Lipid Res.* **60**(2), 323–332 (2019).
13. Fu, J. *et al.* The gut microbiome contributes to a substantial proportion of the variation in blood lipids. *Circ. Res.* **117**(9), 817–824 (2015).
14. Hoen, A. G. *et al.* Association of cesarean delivery and formula supplementation with the stool metabolome of 6-week-old infants. *Metabolites* **11**(10), 702 (2021).
15. Li, N. *et al.* Distinct gut microbiota and metabolite profiles induced by delivery mode in healthy Chinese infants. *J. Proteom.* **10**(232), 104071 (2021).
16. Lu, S., Huang, Q., Wei, B. & Chen, Y. Effects of β-lactam antibiotics on gut microbiota colonization and metabolites in late preterm infants. *Curr. Microbiol.* **77**(12), 3888–3896 (2020).
17. Patton, L. *et al.* Antibiotics effects on the fecal metabolome in preterm infants. *Metabolites* **10**(8), 331 (2020).
18. Brink, L. R. *et al.* Neonatal diet alters fecal microbiota and metabolome profiles at different ages in infants fed breast milk or formula. *Am. J. Clin. Nutr.* **111**, 1190–1202 (2020).
19. He, X. *et al.* Fecal microbiome and metabolome of infants fed bovine MFGM supplemented formula or standard formula with breast-fed infants as reference: a randomized controlled trial. *Sci. Rep.* **9**(1), 11589 (2019).

20. Conta, G. *et al.* Longitudinal multi-omics study of a mother-infant dyad from breastfeeding to weaning: an individualized approach to understand the interactions among diet, fecal metabolome and microbiota composition. *Front. Mol. Biosci.* https://doi.org/10.3389/fmolb.2021.688440 (2021).

21. Patterson, W. B. *et al.* Prenatal exposure to ambient air pollutants and early infant growth and adiposity in the Southern California Mother's Milk Study. *Environ. Health.* **20**(1), 67 (2021).

22. Liu, K. H. *et al.* Microbial metabolite delta-valerobetaine is a diet-dependent obesogen. *Nat Metab.* **3**(12), 1694–1705 (2021).

23. Tian, L. *et al.* Metapone: a bioconductor package for joint pathway testing for untargeted metabolomics data. *Bioinformatics* **38**(14), 3662–3664. https://doi.org/10.1093/bioinformatics/btac364 (2022).

24. Goran, M. I., Walker, R. & Allayee, H. Genetic-related and carbohydrate-related factors affecting liver fat accumulation. *Curr. Opin. Clin. Nutr. Metab. Care.* **15**(4), 392–396 (2012).

25. Schneider, A. L. C., Lazo, M., Selvin, E. & Clark, J. M. Racial differences in nonalcoholic fatty liver disease in the U.S. population. *Obesity* **22**(1), 292–299. https://doi.org/10.1002/oby.20426 (2014).

26. Wandro, S. *et al.* The microbiome and metabolome of preterm infant stool are personalized and not driven by health outcomes, including necrotizing enterocolitis and late-onset sepsis. *mSphere* https://doi.org/10.1128/mSphere.00104-18 (2018).

27. Brewington, C. R., Parks, O. W. & Schwartz, D. P. Conjugated compounds in cow's milk. *J. Agric. Food Chem.* **21**(1), 38–39 (1973).

28. PubChem. Methyl vanillate. Available from: https://pubchem.ncbi.nlm.nih.gov/compound/19844

29. Marszalek-Grabska, M. *et al.* Unexpected content of kynurenine in mother's milk and infant formulas. *Sci. Rep.* **12**(1), 6464 (2022).

30. Salem, N. & Van Dael, P. Arachidonic acid in human milk. *Nutrients* **12**(3), 626 (2020).

31. Horrocks, L. A. & Yeo, Y. K. Health benefits of docosahexaenoic acid (DHA). *Pharmacol. Res.* **40**(3), 211–225 (1999).

32 Guan, H. *et al.* Comparison of fecal collection methods on variation in gut metagenomics and untargeted metabolomics. *mSphere* **6**(5), e00636 (2021).

33. Wang, Z. *et al.* Comparison of fecal collection methods for microbiome and metabolomics studies. *Front. Cell Infect. Microbiol.* **28**(8), 301 (2018).

34. Plows, J. F. *et al.* Longitudinal changes in human milk oligosaccharides (HMOs) over the course of 24 months of lactation. *J. Nutr.* **151**(4), 876–882 (2021).

35. Hollingshead, A. B. Four factor index of social status. *Yale J. Sociol.* **8** (2011).

36. Go, Y. M. *et al.* Reference standardization for mass spectrometry and high-resolution metabolomics applications to exposome research. *Toxicol. Sci.* **148**(2), 531–543 (2015).

37. Liang, D. *et al.* Use of high-resolution metabolomics for the identification of metabolic signals associated with traffic-related air pollution. *Environ. Int.* **120**, 145–154 (2018).

38. Uppal, K. *et al.* xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinformat.* **14**(1), 15 (2013).

39. Yu, T. & Jones, D. P. Improving peak detection in high-resolution LC/MS metabolomics data using preexisting knowledge and machine learning approach. *Bioinformatics* **30**(20), 2941–2948 (2014).

40. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**(3), 211–221 (2007).

41. Chen, H. In *VennDiagram: Generate High-Resolution Venn and Euler Plots.* Available from: https://CRAN.R-project.org/package=VennDiagram (2022).

42. Gu, Z. In: *ComplexHeatmap: Make Complex Heatmaps. Bioconductor version: Release (3.15).* Available from: https://bioconductor.org/packages/ComplexHeatmap/ (2022).

43. Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R. et al. In: *Vegan: Community Ecology Package.* Available from: https://CRAN.R-project.org/package=vegan (2022).

44. Bates, D., Maechler, M., Bolker [aut, B., cre, Walker, S., Christensen, R. H. B. et al. In: *lme4: Linear Mixed-Effects Models Using "Eigen" and S4.* Available from: https://CRAN.R-project.org/package=lme4 (2022)

45. 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**(1), 289–300 (1995).

46. Efron, B. & Tibshirani, R. Empirical bayes methods and false discovery rates for microarrays. *Genet. Epidemiol.* **23**(1), 70–86 (2002).

Acknowledgements

Author contributions

E.A.H., B.C., and T.L.A. had full access to all the data in the study and take responsibility for the integrity of the data and accuracy of the data analysis. E.A.H., N.S., B.C., and Z.L. analyzed the data. V.T. conducted the experiments, H.H.C. provided statistical consultation, J.A.S., D.P.J., M.I.G., and H.H.C. assisted with manuscript preparation and interpretation of results. All authors provided feedback and approved the final manuscript. M.I.G. conceived of the original Mother's Milk Cohort and D.L. and T.L.A. conceived of the study and designed the experiments.

Competing interests

MIG receives book royalties and is a scientific advisor for Yumi. The authors declare no additional competing interests to disclose.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-28862-z.

Correspondence and requests for materials should be addressed to T.L.A.

Reprints and permissions information is available at www.nature.com/reprints.

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