

# Effects of Diet on the Microbiome and Serum Metabolome of South Asian Infants at 1 Year

Colin Y. Bruce<sup>1</sup>, Meera Shanmuganathan<sup>2</sup>, Sandi M. Azab<sup>3</sup>, Philip Britz-McKibbin<sup>2</sup>, Sonia S. Anand<sup>3,4</sup>, Russell J. de Souza<sup>4</sup> and Jennifer C. Stearns<sup>3,5</sup>

<sup>1</sup> Department of Biochemistry; [brucec6@mcmaster.ca](mailto:brucec6@mcmaster.ca) (C.Y.B)

<sup>2</sup> Department of Chemistry and Chemical Biology; [shanmm2@mcmaster.ca](mailto:shanmm2@mcmaster.ca) (M.S); [britz@mcmaster.ca](mailto:britz@mcmaster.ca) (P.B-M)

<sup>3</sup> Department of Medicine; [azabs@mcmaster.ca](mailto:azabs@mcmaster.ca) (S.M.A); [anands@mcmaster.ca](mailto:anands@mcmaster.ca) (S.S.A), [stearns@mcmaster.ca](mailto:stearns@mcmaster.ca) (J.C.S)

<sup>4</sup> Department of Health Research Methods, Evidence & Impact; McMaster University, Hamilton, ON, Canada; [desouzrj@mcmaster.ca](mailto:desouzrj@mcmaster.ca)

<sup>5</sup> Farncombe Family Digestive Health Research Institute

\* Correspondence: [stearns@mcmaster.ca](mailto:stearns@mcmaster.ca)

**Abstract:** Diet is known to affect the gut microbiome and metabolome composition in adults, but this has not been fully explored in infants. Dietary patterns from 1 year-old infants (n=182) from the South Asian Birth Cohort (START) study were compared to gut microbiome alpha and beta diversity and to taxa abundance differences. Diet – serum metabolite associations were identified using multivariate analysis (partial least squares-discriminant analysis, PLS-DA) and univariate analysis (T-Test). Dietary biomarkers identified from START were also examined in a separate cohort of white Caucasian infants ([CHILD Cohort Study](#), n=82). Lastly, the association of diet with gut microbiome and serum biomarkers, considering maternal, perinatal and infant characteristics was investigated using multivariate forward stepwise regression. A dietary pattern characterized by breastfeeding, supplemented by formula and dairy was the strongest predictor of the gut microbiome that also differentiated the serum metabolome of infants. The formula and dairy dietary pattern was associated with a panel of circulating metabolites in both cohorts, including: S-methylcysteine, branched-chain/aromatic amino acids, lysine, dimethylglycine, and methionine. Breastfeeding status, the prominent feature of the dietary pattern, was also associated with a sub-set of serum metabolites in both cohorts. In START, this diet pattern was associated with the metabolites tryptophan betaine, 2-hydroxybutyric acid, tyrosine, phenylalanine, and trimethyl-N-oxide. In the CHILD Cohort Study(CHILD), breastfeeding status was associated with the metabolites aminooctanoic acid, 3-hydroxybutyric acid, and methyl-proline. The results of our study suggest that breastfeeding has the largest effect on the composition of the gut microbiome and the serum metabolome at 1 year, even when solid food diet and other covariates are considered.

**Keywords:** infant diet; gut microbiome; serum metabolome; breastfeeding; ethnicity.

## 1. Introduction

Early life risk factors may contribute to chronic diseases of adulthood [1–3]. Several early life factors have been identified as predictors of future health outcomes. For example, rapid weight gain during early life increases risk of adult cardiovascular disease [3,4]. Infant vegetarianism or plant-based diets may promote immune system regulation through induction of T-regulatory cells and decrease risk of auto-immune disorders [5]. A longer duration of breastfeeding during infancy has been linked to decreasing incidence of gastrointestinal infections, asthma, and obesity [6,7].

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

Such diverse exposures may exert their effects through a common pathway—namely, the

modulation of the gut microbiome [8]. The gut microbiome undergoes significant compositional changes during infancy and is emerging as an important predictor of long-term health [5,9] and chronic disease risk. Over the first year of life significant changes in diversity, richness, and relative abundances of gut microbes are observed as the microbiome matures and approaches an adult-like composition over the first 3-5 years of life [10–13]. A driver of gut microbial maturation is the infant diet, which also changes in composition from early life milk-based diets, towards adult-like solid food based diets [14]. Breastfeeding during early life supplies common milk substrates such as lactose and human milk oligosaccharides (HMOs) to the infant gut [10,14,15]. This substrate availability, especially HMOs, enriches for microbes characteristic of the early infant gut microbiome such as members of the genera *Bifidobacterium*, *Lactobacillus* and *Bacteroides*. These decrease in abundance at weaning [10,12,16]. The gut microbiomes of infants who are predominantly formula-fed exhibit a distinct microbiome composition from those who are breastfed, show lower abundances of *Bifidobacterium* compared to breastfed infants of the same age [15], and mature differently than the gut microbiomes of breastfed-dominant infants [14,17]. While the effects of breastfeeding and formula feeding on the infant gut microbiota have been well investigated, less is known about how the timing and nature of solid foods introduced influences this difference. The transition towards a solid food diet has been shown to increase the complexity of the gut microbiome, however, predominant feeding type may influence these processes [10,17].

Another potential mechanism for breastfeeding and early-life diet to influence health outcomes is via bioactive molecules mediating complex host-microbiome interactions. These metabolites may originate directly from specific foods in the diet [18], comprising both nutrient and non-nutrient food components where they can be classified as either *exposure biomarkers*, in which the food components exert some physiological effect on the host, or as *outcome biomarkers* following host and/or gut microbiota biotransformations [19]. A growing field of nutritional metabolomics investigates the relationship between diet and serum or urinary metabolites that may serve as objective *biomarkers of food intake* (BFIs) [20–23]. Alternatively, metabolites may be microbially-produced and secreted into the gut environment, such as trimethylamine-*N*-oxide (TMAO), which is derived from intake of fish [24] and/or synthesized via microbial metabolism of dietary carnitine or phosphatidylcholine [25]. Common substrates for microbial metabolism from dietary fibres also include short-chain fatty acids (SCFAs), notably acetate, butyrate, and propionate [26,27].

In this study, we investigated the association of host diet with the gut microbiome, diet with the serum metabolome, and the gut microbiome with the serum metabolome during the first year of life in a cohort of South Asian infants residing in the Greater Toronto Area (Ontario), which was also compared with a subset of white Caucasian infants from a pan-Canadian cohort.

## 2. Materials and Methods

### Study Design/participants :

The South Asian Birth Cohort (START-Canada) enrolled 1012 South Asian mother–child pairs from the Peel Region of Ontario to investigate the influence of diverse environmental exposures and genetics on early life adiposity, growth trajectory, and cardiometabolic risk factors [28]. The infant was considered to be of South Asian ethnicity if both parents and grandparents' ancestral origin was identified as from India, Pakistan, Sri Lanka, or Bangladesh. Participant data,

including dietary data and microbiome data, for 182 samples were collected [8]. Serum metabolite data for 72 participants were also collected from START. For comparison of results across ethnicities, and for replication of the serum metabolite analysis, participant data (diet, microbiome, serum metabolite) for 82 participants that reported white Caucasian as their ethnicity was collected from CHILD [29]. CHILD participants included in this study were recruited across multiple centres in Canada (Toronto (n=21), Winnipeg (n=1), Edmonton (n=30), Vancouver (n=29)).

#### Dietary data collection

Maternal and infant diet data were collected using validated semi-quantitative food frequency questionnaires (FFQs) [30–33]. The maternal semi-quantitative FFQ was developed and validated for use among South Asians living in Canada. On the FFQ, participants report how often, on average, they consumed selected foods in the previous year. Maternal diet data was collected during the second trimester of pregnancy. From the FFQ, we created 36 food groups as described previously [20,34], which we used to develop a diet quality score (DQS), calculated as the sum of daily servings of “healthy” foods (fermented dairy, fish and seafood, leafy green vegetables, cruciferous vegetables, legumes, fruits, nuts, and whole grains) less the sum of daily servings of “unhealthy foods” (processed meats, refined grains, French fries, snacks, sweets, and sweet drinks. These foods were chosen because they have been widely used to characterize healthy dietary patterns (i.e., prudent diet) that reduce chronic disease risk [35–42]. Our DQS correlates well with a modified version of the AHEI previously derived in these cohorts [20,34,43]. The infant semi-quantitative FFQ built upon existing instruments used in 3 studies: the Avon Longitudinal Study of Parents and Children (ALSPAC) [44], the Born in Bradford (BiB) study [45], and the Study of Health Assessment and Risk Evaluation (SHARE) [30]. These tools were developed for use in children, South Asian children, and South Asians living in Canada, respectively. Starting with items listed on any of the 3 instruments, dietitians and study personnel of South Asian origin refined the list to add missing foods, remove redundant foods, and produce a booklet that queried the infant’s frequency of consumption of 80 individual food items over the preceding 4 weeks. Respondents were asked to quantify the number of “servings/day”, “servings/week”, or “servings/month” as appropriate. Reference portion sizes were provided for each item, guided by the serving sizes in the ALSPAC instrument. The 80 food items were then sorted into common food groups resulting in 12 food groups: Refined Grains, Whole Grains, Processed Meats, Unprocessed Meats, Fish, Vegetables, Beans and legumes/nuts/plant proteins, Sweets, Fruits, Dairy, Eggs, and Other.

The infant diet questionnaire used for the CHILD cohort was similar to the one used with START. The CHILD infant diet instrument used a 13-item questionnaire asking about consumption of groups of foods that had been introduced in the first year of life. The 13 food items listed on the questionnaire were: Prepared baby food, Baby cereals, Commercial cereals, Grains, Dairy products, Vegetables, Fruits, Meats, Egg, Fish, Shellfish, Nuts, Peanuts/peanut butter.

#### Health data collection

Health and other participant characteristic data included: maternal vegetarian status, weight gain over the first year of life, age in weeks at the 1 year sample collection, breastfeeding status at

1 year, current formula feeding status at 1 year, and offspring age in weeks when solid foods were introduced [1,34].

#### Gut microbiome stool analysis:

Sequence reads for the v3 region of the 16S rRNA genes from 182 infant stool samples from our previous study [8] were reprocessed for this study. Adapter, primer and barcode sequences were trimmed from sequencing reads with cutadapt [46], then amplicon sequence variants (ASV) were inferred from the sequenced data using DADA2 [47]. Taxonomy was assigned in DADA2 that uses the RDP naive Bayesian classifier method using the SILVA 16S rRNA gene reference file [48] (release version 132). Rarefied species richness [49] ACE, Chao1 and the inverse Simpson [50] and Shannon diversity were calculated from ASV count data with the vegan [51] package in R. Between-community diversity analyses were based on the Aitchison distance [52] calculated using the Euclidean distance of centre-log transformed ASV level count data. A PCA of Aitchison distance was calculated with x to generate an eigen value for each sample from PC1 and PC2. Bacteroidetes:Firmicutes ratios were calculated for each sample by agglomerating taxa at the phylum level for each sample and dividing the total Bacteroidetes read counts for each sample by the total Firmicutes read counts. The same process was done for calculating the Firmicutes:Actinobacteria ratio.

#### Serum sample pre-treatment and instrumental analysis for serum metabolomics:

Frozen serum samples (START (n=72), CHILD (n=82)) were thawed on ice and an aliquot of 50  $\mu$ L was transferred into a 0.5 mL centrifuge tubes for sample dilution (a 4-fold) with internal/recovery standards followed by ultrafiltration (3 kDa) to remove protein (i.e., serum filtrate) as described previously [53]. In addition, a pooled quality control (QC) was prepared by combining an equal aliquot of serum samples from START maternal cohort (n=300) that was used to assess technical precision (This was the first cohort to be analyzed in the lab). All serum filtrate and QC samples were stored frozen at -80°C prior to metabolomic analysis by multisegment injection-capillary electrophoresis-mass spectrometry (MSI-CE-MS). A standardized protocol was used to measure polar/ionic serum metabolites under positive and negative ion mode detection with full-scan data acquisition [53]. This technique takes advantage of a serial sample injection format of seven (or more) samples within a single run [54] using an Agilent 6230 time-of-flight mass spectrometer (TOF-MS) with a coaxial sheath liquid Jetstream electrospray ion source coupled to an Agilent G7100A capillary electrophoresis (CE) unit.

Following data acquisition, all raw MSI-CE-MS data (.d format) were processed using MassHunter Workstation Qualitative Analysis software (version B.06.00, Agilent Technologies, 2012). Both a targeted (i.e., known metabolites confirmed with reference standards) and a nontargeted (i.e., unknown ions authenticated from molecular features and putatively identified by MS/MS) workflows were used to characterize the infant serum metabolome [53], where all ions were annotated based on their accurate mass-to-charge ratio (m/z), relative migration time (RMT) and mode of detection (p:positive, n:negative). Peak areas and migration times for all serum metabolites and internal/recovery standards were transferred to an Excel worksheet (Microsoft Office) and relative peak areas (RPA) for each unique molecular feature were exported as a .csv file.

Serum metabolites frequently detected in > 75% of all samples analyzed with adequate technical precision (CV < 30% for repeat QC samples analyzed) were included in the data matrix for further statistical analysis. Missing value inputs were replaced by a value equal to one-half the lower limit of detection. Serum metabolite concentrations ( $\mu\text{M}$ ) were reported for all known/confirmed serum metabolites, whereas relative peak area (RPA) were used for tentatively identified or unknown metabolites lacking available standards.

### Statistical analysis

Linear regression models were fitted to test the association of diet patterns and participant characteristics with gut microbiome alpha diversity. Univariable tests of the association of diet pattern, and participant characteristics with gut microbiome beta diversity were done with PERMANOVA with 99,999 permutations using the vegan package in R. Multivariable models to assess the effect of diet patterns and participant characteristics on gut microbiome beta diversity were fitted with the envfit function in the vegan package. Bacterial ASV enrichment based on the formula and dairy dietary pattern were measured with the DESeq2 package in R [55].

To assess the relationship between diet and the serum metabolome, diet patterns were compared to serum metabolite concentrations using the Metaboanalyst [56] web platform. Serum metabolite data were *log*-transformed and auto-scaled prior to supervised multivariate data analysis using partial least squares-discriminant analysis (PLS-DA). A variable importance in projection (VIP) plot was used to rank order serum metabolites of significance to the PLS-DA model. Metabolites that had the greatest contribution to the differences between groups in the PLS-DA model were carried forward for further investigation. For START samples univariate testing (Tukey's T-test, FDR adjusted for multiple hypothesis testing) was used to confirm the significance of serum metabolites with diet patterns and an adjusted  $p < 0.05$  was considered significant. Univariate metabolite analysis was adjusted for covariates listed in Table S4. A similar analysis was done to compare serum metabolites associated with breastfeeding status in START participants at 1-year collection time. To confirm the associations of serum metabolites with breastfeeding status the same analysis was repeated using samples from CHILD. The association of diet and participant characteristics with the concentration of serum metabolites of interest was modeled with univariate and multivariate models that included a combined dataset from START and CHILD. Variables significant at  $p < 0.1$  in univariate testing were carried into a forward stepwise regression model.

### Self-reported dietary intake:

We took a targeted approach to assessing the association between infant food intake and circulating metabolite concentrations in non-fasting serum. We started with a list of previously identified robust dietary biomarkers of specific foods in adults [20,24] within this infant population. These included proline betaine (citrus fruits), methylhistidine (red meat, chicken, eggs), hippuric acid (vegetable, fruit), TMAO (seafood, meat, red meat, eggs), and tryptophan betaine (nuts, seeds, peanuts, legumes).

We explored several approaches to characterizing the diet in this analysis, as follows: 1) We created a diet diversity index ("dietary diversity 1"), which was a simple count of the number of different food groups (i.e. out of 36 for mothers; out of 12 for infants) that an individual reported

consuming a non-zero amount of; 2) We created a diet richness index (“dietary diversity 2”) in which we simply counted the number of items (i.e. out of 152 for mothers; out of 80 for infants) that were consumed a non-zero amount of times); 3) We used principal components analysis (PCA) to derive dietary pattern scores for mothers and infants (as described previously [34]); 4) We used a PCA to derive 3 dietary patterns for infants, using the 12 food groups; these patterns were called animal foods, plant foods, and infant formula and dairy.

### 3. Results

#### Description of the infant diet:

The interquartile range for diet diversity 1 was 7 to 9; and for diet diversity 2, 2 to 3. Three distinct dietary patterns were retained from the unsupervised PCA model, 1: Animal foods (unprocessed meat, processed meat, fish, eggs; eigenvalue = 1.43), 2: Plant foods (whole grains, vegetables, beans/legumes, fruits, tea/water; eigenvalue = 1.16), 3: Formula/dairy (dairy, formula, with a negative loading on breastfeeding; eigenvalue = 0.92), which together explained 44.6% of the variance in diet.

#### Diet and participant characteristics associated with infant gut microbiome composition:

We investigated the association between infant variables and gut microbial alpha diversity from START participants and found that infant weight gain over the first year of life was negatively associated with Shannon diversity ( $-0.063$ ,  $p = 0.009$ ), however, no association was found between diet diversity or the diet patterns and microbial alpha diversity estimates. Comparison of the same set of diet variables and participant characteristics to beta diversity (Aitchison distance) identified that number of items eaten during the reporting period and age in months at 1 year collection time were associated with beta diversity in univariable PERMANOVA analysis (Table 1). The multivariable *envfit* model indicated that adherence to the formula and dairy dietary pattern and age were associated with beta diversity (Table 1). For *envfit* analysis the breastfeeding status was not included in the same model with adherence to the formula and dairy dietary pattern because both variables include breastfeeding as a component, and thus were highly collinear ( $r = -0.77$ ). In fact, low adherence to the formula and dairy dietary pattern corresponded to current breastfeeding status, and high adherence corresponded to 2.22 servings of formula and 6.67 serving of dairy.

**Table 1.** Beta Diversity Scores Associated with Dietary Variables

253

Variable	Univariable R <sup>2</sup> (P-Value)	Multivariable R <sup>2</sup> (P-value)
Dietary Diversity	0.006 (0.166)	0.034 (0.754)
Number of Items Eaten	<b>0.008 (0.009)</b>	0.026 (0.519)
Adherence to Animal Foods DP	0.006 (0.243)	0.003(0.930)
Adherence to Plant Foods DP	0.006 (0.229)	0.013(0.350)
Adherence to Formula and Dairy DP	<b>0.007 (0.078)</b>	<b>0.109 (&lt;0.001)</b>
Breastfeeding status	<b>0.008 (0.004)</b>	
Mother Vegetarian Status in Pregnancy	0.006 (0.182)	0.017(0.189)
weight gain in the first year	0.007 (0.062)	0.007(0.536)
Age in months at collection	<b>0.008 (0.014)</b>	<b>0.046 (0.015)</b>
Age when solid foods introduced	0.006 (0.381)	0.031(0.197)

DP = dietary pattern

254

255

We next looked at the the enrichment of gut microbial taxa based on adherence to the formula and dairy dietary pattern and found that many ASV from *Lactobacillus*, *Bifidobacterium*, *Veillonella* and *Megasphaera* were less abundant in infants with a high compared with a low adherence to the formula and dairy dietary pattern (Table S1). We investigated whether these associations were driven by solid food consumption differences between breastfed and not-breastfed infants but we did not find significant differences in the amounts of commonly consumed foods such as legumes (difference = 0.07 servings/d, p-value = 0.34), and eggs (difference = -0.12 servings/d, p-value = 0.18). We also found an enrichment in high adherence to the formula and dairy dietary pattern of ASV from Enterobacteriaceae, *Enterobacter*, *Lactococcus*, *Blautia*, *Erysipelatoclostridium* (Table S1).

256

257

258

259

260

261

262

263

264

265

266

#### Associations of the Infant Diet with the Infant Serum Metabolome:

267

To observe how diet at 1 year may be influencing the serum metabolome, a supervised PLS-DA model was applied to compare differences in the serum metabolic phenotype of infants as a function of their dietary pattern adherence score. We included serum samples from infants with low adherence (n = 26), medium adherence (n= 17), or high adherence (n= 29) to the formula and dairy dietary pattern, and observed almost complete overlap of profiles from medium and high adherence groups, but a distinct metabolome profile for low adherence infants (Figure 1A).

268

269

270

271

272

273

Separation of infant serum metabolite profiles was primarily due to dimethylglycine (DMG), valine, and tryptophan betaine (VIP > 1.5; Figure S1). Univariate (Tukey's T-test, FDR adj. p < 0.05) tests confirmed the associations of these metabolites of interest with adherence to the formula and dairy diet pattern, as well as phenylalanine, S-methylcysteine, leucine, tyrosine, 2-hydroxybutyric acid, and lysine (Table S2). Low adherence to the formula and dairy dietary pattern at 1 year was associated with higher concentrations of tryptophan betaine and S-methylcysteine, whereas medium and high adherence to the formula and dairy dietary pattern was associated with higher lysine, 2-hydroxybutyric acid, dimethylglycine, methionine, TMAO, tyrosine and phenylalanine (aromatic amino acids), leucine and valine (branched-chain amino acids; BCAA)

274

275

276

277

278

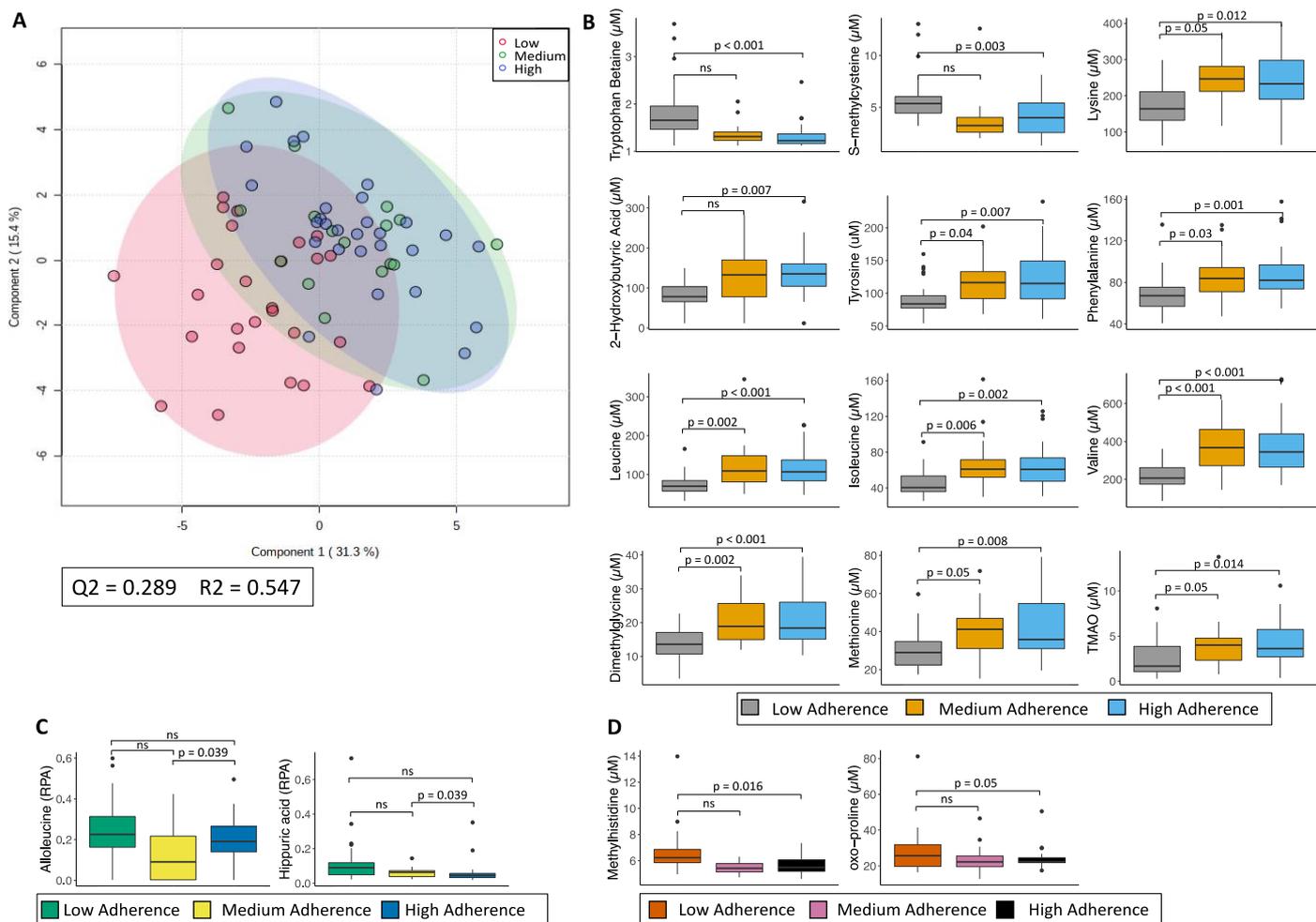
279

280

281

282

(Figure 1B). Tryptophan betaine is a serum biomarker of legume [20,57] and nut intake [24,58], and legume consumption was high in the START cohort [34]. We found no association between circulating tryptophan betaine in infant serum and either maternal or infant legume and nut intake (data not shown).

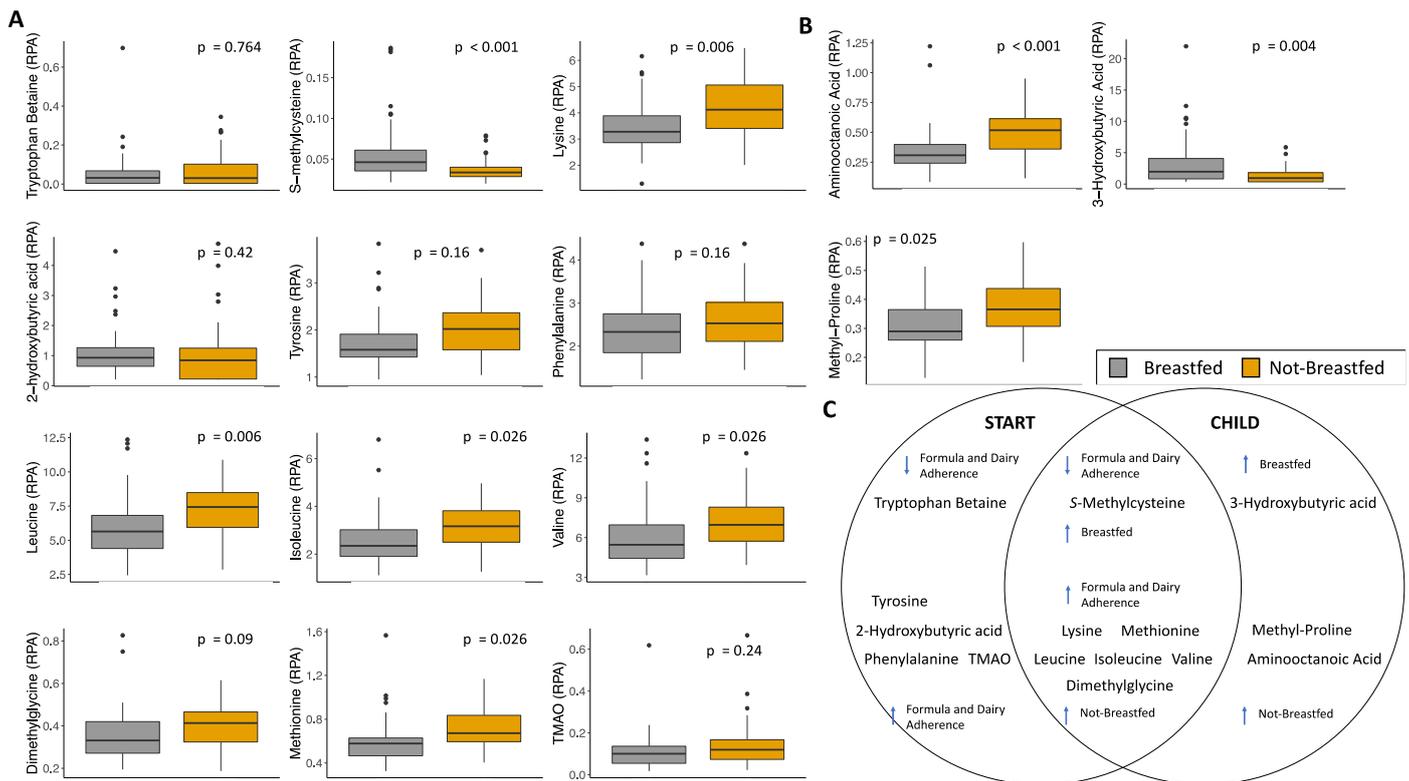


**Figure 1.** Metabolite concentrations related to dietary pattern adherence in the START cohort. **A)** PLS-DA of samples related to formula and dairy dietary pattern adherence with 95% confidence interval. Cross-validation was performed using leave-out-one-at-a-time cross-validation, and Q2 was maximized with 2 principal components. Serum metabolites related to **B)** formula and dairy dietary pattern adherence, **C)** animal foods dietary pattern adherence and **D)** plant foods dietary pattern adherence. Boxplots are coloured by adherence to specific diet pattern at collection time (box = 1st and 3rd quartiles with median, whiskers = 1.5x IQR, dots = outliers). P-values are from Tukey's T-tests, FDR adjusted. For the animal and plant dietary patterns, samples from infant who were breastfeeding status were removed prior to analysis. RPA: relative peak area

We next excluded infants who were breastfeeding at the time of the sample collection from the analysis and compared the serum metabolomic profile based on the adherence to the plant foods

303 dietary pattern and the animal foods dietary pattern. We used univariate (Tukeys' T-test, FDR adj. 303  
304  $p < 0.05$ ) tests on any metabolite with a VIP score  $\geq 1.5$  and found that high adherence to the plant 304  
305 foods dietary pattern was associated with higher alloleucine and hippuric acid than medium 305  
306 adherence, but was not significantly different from low adherence (Figure 1C). Adherence to the 306  
307 animal foods dietary pattern was associated with a lower concentration of methylhistidine and oxo- 307  
308 proline (Figure 1D). Since low adherence to the formula and dairy dietary pattern in START was 308  
309 associated with a greater likelihood of breastfeeding, we looked at the association of serum 309  
310 metabolites with breastfeeding in START. The same approach was used comparing START 310  
311 participants that were breastfed at collection time ( $n = 31$ ) to non-breastfed at collection time ( $n =$  311  
312 41) that generated a PLS-DA model similar to the formula and dairy diet pattern (Figure S2). The 312  
313 breastfed group overlaps closely with the low adherence to the formula and dairy diet pattern group, 313  
314 and the non-breastfed group with the medium and high adherence groups. As expected, the same 314  
315 serum metabolites were determined to differentiate infant metabolite profiles (tryptophan betaine, 315  
316 dimethyl-glycine, valine, VIP  $> 1.5$ ; Figure S2) as a function of breastfeeding adherence patterns. 316  
317 The two groups shared the same top 11 metabolites as ranked by VIP scores (VIP  $> 1.1$ ), with only 317  
318 TMAO not as strongly associated with breastfeeding status as compared to the formula and dairy 318  
319 diet pattern. 319

320 To determine whether serum metabolic phenotypes associated with breastfeeding status that 320  
321 had been observed were specific to START, the same analysis was repeated in an independent 321  
322 cohort of white Caucasian 1-year old infants from CHILD ( $n=117$ ) having otherwise similar 322  
323 characteristics (Table S3) as the infants in START. The results from START were then replicated 323  
324 in CHILD for 7 of 11 serum metabolites (Tukey's T-test, adj.  $p < 0.05$ ). The concentration of *S*- 324  
325 methylcysteine was higher in breastfeeding infants than non-breastfeeding infants, whereas leucine, 325  
326 isoleucine, valine, lysine, dimethylglycine, and circulating methionine was lower (Figure 2A). We 326  
327 then investigated additional association of serum metabolites with breastfeeding status in CHILD 327  
328 alone, in order to identify metabolites unique to this population, which revealed that serum 3- 328  
329 hydroxybutyric acid was higher in breastfed infants, whereas aminooctanoic acid and *N*- 329  
330 methylproline were lower in white Caucasian infants not breastfeeding (Figure 2B). 330



**Figure 2.** Metabolite concentrations within the CHILd cohort. **A)** Serum metabolites investigated since they were associated with the formula and dairy dietary pattern in START. **B)** Serum metabolites uniquely associated with breastfeeding status in CHILd participants. Boxplots of serum metabolite concentrations are coloured by participant breastfeeding status at sample collection time (box = 1st and 3rd quartiles with median, whiskers = 1.5x IQR, dots = outliers). P-values are from Tukey's T-tests, FDR adjusted. **C)** Venn diagram of serum metabolites associated with diet in each study cohort. Arrows indicate the directionality of the association of metabolite concentrations with diet.

Triangulating the associations between infant diet, gut microbiome and serum metabolome:

Having found associations between the infant diet and the gut microbiome and associations between the infant diet and the serum metabolome, we then explored the contribution of diet and the gut microbiome to the concentration of specific serum metabolites while accounting for maternal, perinatal and infant characteristics. We tested whether serum concentration levels of metabolites significantly associated with diet in both cohorts: S-methylcysteine, leucine, isoleucine, valine, lysine, methionine, and dimethylglycine, were explained by feeding-related factors (breastfeeding status at 1 year, time since weaning, formula feeding status at 1 year, age of solid food introduction and cows milk consumption), the gut microbiome (PC1, PC2, Bacteroidetes/Firmicutes ratio and Firmicutes/Actinobacteria ratio), maternal characteristics (years in Canada), perinatal variables (GDM, delivery mode, maternal antibiotic and vegetarian status during pregnancy), or infant characteristics (birth weight, ethnicity, sex, weight gain in the first year and age) using forward stepwise regression (Table S4). Breastfeeding status was associated with all metabolites investigated following forward stepwise regression. S-methylcysteine, valine, leucine, and isoleucine were most

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

strongly associated with breastfeeding status. *S*-methylcysteine concentration was only associated with breastfeeding status, leucine and valine concentrations were associated with breastfeeding status and microbiome PC2, and isoleucine was associated with breastfeeding status and maternal antibiotic use during pregnancy (Table S4). Lysine, methionine, and DMG were most strongly associated with microbiome PC2, followed by breastfeeding status. Lysine concentrations were also associated with maternal vegetarian status, and DMG was associated with formula feeding status at 1 year collection time.

#### 4. Discussion

In this study, we have explored the complex interactions of infant diet in relation to infant serum metabolome and the infant gut microbiome in order to better understand the role of gut microbial communities in infant nutrition and metabolism within a Canadian cohort infants of South Asian and European descent. The results of this study indicate that gut microbiome diversity of 1-year old infants is affected by breastfeeding status at 1 year, but not by solid food consumption. The infant serum metabolome was influenced by host diet, with distinct serum metabolites identified with relation to consumption of animal foods, plant foods, and most strongly with infant formula and dairy foods.

The formula and dairy dietary pattern was associated with a panel of circulating metabolites in both cohorts, including: *S*-methylcysteine, branched-chain/aromatic amino acids, lysine, dimethylglycine, and methionine. The results of our study suggest that breastfeeding has the largest effect on the composition of the gut microbiome and the serum metabolome at 1 year, even when solid food diet and other covariates are considered.

Our multivariate analysis results find that breastfeeding status and infant formula and dairy consumption are the drivers of gut microbiome composition, regardless of the types of solid foods consumed in the first year of life. These results are supported by specific taxa enrichment in response to dietary variables in which microbes associated with breastfeeding, largely belonging to the genera *Bifidobacterium*, *Lactobacillus* and *Veillonella* were consistently less abundant with increasing formula and dairy diet pattern adherence. The lower abundance of *Bifidobacterium* and *Lactobacillus* in response to cessation of breastfeeding is likely due to altered substrate availability, and is consistent with previous studies that showed similar decreases in both genera [10]. Increasing formula and dairy consumption was associated with a higher abundance of the genus *Blautia*, which are characteristic of late infancy and adult populations [10]. This increase is likely in response to the transition away from breastfeeding because *Blautia* are unable to metabolize HMOs [59][60].

Infant age was also a predictor of microbial  $\beta$  diversity. Age has been previously considered to be an important factor in the maturation of the gut microbiome [8,10,17]. Previously, infant formula feeding has shown to produce a more adult-like gut microbiome than breastfeeding [10]. Weaning from breastmilk and the transition to solid foods has also been suggested to be a driver of gut microbial change [10].

We identified tryptophan betaine and *S*-methylcysteine as dietary biomarkers associated with breastfeeding status among infants, with tryptophan betaine uniquely associated with breastfeeding in South Asian participants. Previous studies have detected tryptophan betaine in breastmilk samples [61], suggesting its transfer into the infant gut through breastfeeding. Future work may include

testing samples of breastmilk for the presence of these metabolites. Tryptophan betaine, also known as hypaphorine, is an indole alkaloid that has been previously measured in nuts and legumes (e.g., chickpeas and lentils) [20,24,57]. A small study in breastfeeding mothers (n =24) reported its presence in breast milk and also an association with peanut consumption [61]. Dietary challenge in lactating women with hypaphorine-rich foods (restricted intake for 3 days before the intervention) demonstrated transfer of hypaphorine into milk with hypaphorine appearance peaking 5-18 h after consumption. Prolonged disappearance of hypaphorine from the milk was indicative of slow excretion or metabolism (over 2-3 days) [61]. This suggests that there may be an ideal window to capture BFIs in the infants serum. We previously reported a strong association between nut and legume consumption and measured serum tryptophan betaine concentrations in pregnant women from START (n=300) [20]. However, in the present analyses, we did not detect an association of maternal nut and legume intake with infant serum tryptophan betaine (0.007, p = 0.64). The biomarkers presented here reflect cultural dietary practices, such as vegetarianism and habitual intake of legume-rich foods among the South Asians in our study. Infant nursing practices may also play a part in these differences, as START infants were more commonly formula fed or mixed-fed at collection time compared to white Caucasian infants from CHLD [8]. Identifying dietary biomarkers may allow a more tailored approach for investigating infant nutrition and the potential sources of nutrients, as they may originate from maternal or infant diets.

*S*-methylcysteine was shown to be strongly associated with breastfeeding status among both cohorts, independently of ethnicity, suggesting a potential biomarker of breastfeeding. *S*-methylcysteine has previously been reported to increase in host serum following consumption of navy beans, garlic, and onions [62–64]. *S*-methylcysteine may be related to host health benefits including reducing risk of cancer, and potentially anti-inflammatory activity [62,65]. Previous reports have not linked *S*-methylcysteine concentrations to breastfeeding. It is possible that this metabolite behaves similarly to tryptophan betaine, introduced through consumption of beans and passed to the infant through breastmilk. Correlational analysis between *S*-methylcysteine and other serum metabolites identified moderate correlations with glycine (r = 0.44), glutamine (r = 0.41), tryptophan betaine (0.40), suggesting these as potential co-occurring metabolites, as well as weak correlations with serine (r = 0.34) and choline (r = 0.30). Glycine has been reported to be common in vegetarian and vegan diets, and is commonly consumed in soy products [66,67]. Glutamine is an amino acid commonly found in meat products, but may also be found in plant foods including soy and peanuts [68]. Tryptophan betaine, as we have previously described, is commonly found in nuts and legumes [20,24,57]. Serine is a precursor for *S*-methylcysteine in beans, and acts as a precursor for glycine production in humans [69,70]. Choline is commonly found in animal products, but is also found in beans and soy products [62]. The sources of metabolites that are positively correlated with *S*-methylcysteine concentration suggest the source of this metabolite to be from beans and bean product consumption. The relationship of *S*-methylcysteine with breastfeeding is still unclear, but may represent a mode of transfer to the infant.

Serum metabolites that were associated with adherence to the plant and animal foods diets have been previously identified in foods related to those patterns. Hippuric acid, identified to be in higher concentration in participants who consumed a high plant foods adherent diet, has previously

been shown to be prevalent in a variety of plant foods [20,22,71]. Similarly, methylhistidine is found in a number of animal products and is related to protein turn-over [20,21,24,72–74].

The serum metabolites that were inversely associated with breastfeeding status in the START cohort included dimethylglycine, valine, phenylalanine, leucine, isoleucine tyrosine, lysine, methionine, TMAO, and 2-hydroxybutyric acid. The same associations were replicated in CHILD for the metabolites valine, leucine, isoleucine, lysine, and methionine.

Two metabolites were inversely associated with breastfeeding status and were unique to the CHILD cohort, aminooctanoic acid and 3-hydroxybutyric acid. 3-hydroxybutyric acid (3-hba) is a ketone body, commonly associated with exercise and periods of fasting [75]. 3-hba may be a marker of type II diabetes, when it is significantly higher in plasma compared to control groups [76]. It may also have a role in repressing immune activated inflammation related to obesity and aging [77].

The identification of the BCAAs leucine, isoleucine, and valine, may allow them to serve as biomarkers of formula consumption in infants. The prevalence of BCAAs and TMAO in the serum of formula fed infants may be due to higher protein and choline concentration in infant formulas compared to breastmilk [78–80], especially breastmilk produced when the infant is 1 year of age.

The infant gut microbiome, particularly the presence of *Bifidobacterium* during early life, has been associated with development of the immune system, prevention of allergy, and the production of nutrients available to the host [10,81]. The identification of objective dietary biomarkers for breastfeeding, formula feeding and the solid food diets will enhance our understanding of the contribution of the infant diet to metabolism and development. Future studies should investigate how dietary biomarkers are introduced into the infant serum, and whether they may be passed from mother to offspring during breastfeeding, as well as their effects on offspring health.

#### Strengths and limitations:

This study investigated a unique population of infants, the South Asian population, that has dietary characteristics distinct from white Caucasian populations, for instance the prevalence of vegetarianism and the higher rates of formula and mixed feeding. A strength of this study is the validation of diet-metabolite associations in a second cohort of white Caucasian infants with the identification of ethnic-specific metabolites. Importantly, we replicate the association of dietary biomarkers previously validated in the adult population in infants. Potential weaknesses of this study are the small sample sizes used for the metabolite analysis, as only 153 infants in this study had serum samples collected, which were collected under non-fasting conditions. In addition, only polar ionic metabolites were analyzed. Future work should be expanded to lipid classes of metabolites, and to include analysis of lipids from serum extracts as dietary biomarkers of full-fat dairy intake [82].

#### **5. Conclusions**

In this cohort of 72 South Asian infants aged 1 year, breastfeeding status has the largest effect of all dietary variables on gut microbiome composition. Infant formula consumption, dairy consumption, and other animal food consumption lead to a decrease in gut microbes characteristic of the healthy infant gut microbiome, primarily *Bifidobacterium*. Plant based diets may prevent the decrease in *Bifidobacterium* seen with other solid food diets. Serum tryptophan betaine was associated with breastfeeding status only in South Asian infants , whereas serum S-methylcysteine

and branched-chain amino acids were associated with breastfeeding status and formula feeding status respectively, across two cohorts, and may serve as generalizable dietary biomarkers for infants.

**Author Contributions:** Conceptualization, S.S.A and J.C.S; methodology, R.J.d.S, P.B-M, and J.C.S; formal analysis, C.Y.B; resources, P.B-M and J.C.S; data curation, M.S, P.B-M, and R.J.d.S.; writing—original draft preparation, C.Y.B and J.C.S; writing—review and editing, All; visualization, C.Y.B; supervision, SSA; project administration, S.S.A; funding acquisition, S.S.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by a Canadian Institute for Health Research (CIHR) Grant in Food & Health Population Health Research grant (RFA# 201301FH6; 2013-2018). START data were collected as part of a bilateral ICMR /CIHR funded programme (INC-109205) and HSF Canada Grant in Aid (NA7283). S.S.A. holds a Canada Research Chair Tier 1 in Ethnic Diversity and Cardiovascular Disease and a Heart and Stroke Foundation / Michael G. De Groote Chair in Population Health Research. Canadian Institutes of Health Research (S.S.A.), Genome Canada (P.B.-M.) and Canada Foundation for Innovation (P.B.-M.), and J.C.S is supported by the Farncombe Chair in Microbial Ecology and Bioinformatics.

**START study Team Acknowledgements:** South Asian Birth Cohort (START) study data were collected as part of a bilateral program funded by the Indian Council of Medical Research/Canadian Institutes of Health Research (grant INC-109205) as well as additional funding from Heart and Stroke Foundation of Canada grant NA7283. Dr Sonia Anand and Dr. Milan Gupta were the founding co-PIs for START. Sonia Anand is supported by a Tier 1 Canada Research Chair in Ethnic Diversity and Cardiovascular Disease, and a Heart and Stroke Foundation/Michael G. DeGroote Chair in Populaiton Health Research at McMaster University.

**Data Availability Statement:** The datasets generated and/or analyzed during the current study are not publicly available, since the CHILD and START studies are bound by consent and cannot provide identifiable information to an outside group, but are available from the corresponding author on reasonable request.

**Acknowledgments:** We thank the CHILD Cohort Study (CHILD) participant families for their dedication and commitment to advancing health research. CHILD was initially funded by CIHR and AllerGen NCE. Visit CHILD at [childcohort.ca](http://childcohort.ca)

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Zulyniak, M.A.; Souza, R.J. De; Shaikh, M.; Desai, D.; Lefebvre, D.L.; Gupta, M.; Wilson, J.; Wahi, G.; Subbarao, P.; Becker, A.B.; et al. Does the impact of a plant-based diet during pregnancy on birth weight differ by ethnicity? A dietary pattern analysis from a prospective Canadian birth cohort alliance. *BMJ Open* **2017**, doi:10.1136/bmjopen-2017-017753.
2. Nordman, H.; Voutilainen, R.; Laitinen, T.; Antikainen, L.; Huopio, H.; Heinonen, S.; Jaaskelainen, J. Growth and Cardiovascular Risk Factors in Prepubertal Children Born Large or Small for Gestational Age. *Horm. Res. Paediatr.* **2016**, *85*, 11–17, doi:10.1159/000441652.
3. Barker, D.J.P. The developmental origins of chronic adult disease. *Acta Paediatr.* **2004**, *26*–33, doi:10.1080/08035320410022730.
4. Kjaer, T.W.; Medrano, R.; Wojcicki, J.M. Rapid infant weight gain and prenatal maternal depression are associated with prehypertension / hypertension before age 9. *J. Hum. Hypertens.* **2020**, *795*–801, doi:10.1038/s41371-020-0306-8.
5. Haase, S.; Haghikia, A.; Wilck, N.; Muller, D.N.; Linker, R.A. Impacts of microbiome metabolites on immune regulation and autoimmunity. *Immunology* **2018**, *154*, 230–238, doi:10.1111/imm.12933.
6. Jwa, S.C.; Fujiwara, T.; Kondo, N. Latent Protective Effects of Breastfeeding on Late Childhood Overweight and Obesity : A Nationwide Prospective Study. *Obesity* **2014**, *22*, 1527–1537, doi:10.1002/oby.20735.

7. Barbara, R.; Kohlhuber, M.; Schwegler, U.; Fromme, H.; Abou-dakn, M. Breastfeeding duration and exclusivity associated with infants' health and growth: data from a prospective cohort study in Bavaria, Germany. *Acta Paediatr.* **2009**, 974–980, doi:10.1111/j.1651-2227.2009.01281.x. 526  
527  
528
8. Stearns, J.C.; Zulyniak, M.A.; Souza, R.J. De; Campbell, N.C.; Fontes, M.; Shaikh, M.; Sears, M.R.; Becker, A.B.; Mandhane, P.J.; Subbarao, P.; et al. Ethnic and diet-related differences in the healthy infant microbiome. *Genome Med.* **2017**, 1–12, doi:10.1186/s13073-017-0421-5. 529  
530  
531
9. Makki, K.; Deehan, E.C.; Walter, J.; Backhed, F. The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. *Cell Host Microbe* **2018**, 23, 705–715, doi:10.1016/j.chom.2018.05.012. 532  
533
10. Backhed, F.; Roswall, J.; Peng, Y.; Feng, Q.; Jia, H.; Kovatcheva-datchary, P. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe* **2015**, 17, 690–703, doi:10.1016/j.chom.2015.04.004. 534  
535
11. Yatsunenko, T.; Rey, F.E.; Manary, M.J.; Trehan, I.; Dominguez-Bello, M.G.; Contreras, M.; Magris, M.; Hidalgo, G.; Baldassano, R.N.; Anokhin, A.P.; et al. Human gut microbiome viewed across age and geography. *Nature* 2012, 486, 222–227. 536  
537
12. Baumann-Dudenhoefler, A.M.; D'Souza, A.W.; Tarr, P.I.; Warner, B.B.; Dantas, G. Infant diet and maternal gestational weight gain predict early metabolic maturation of gut microbiomes. *Nat. Med.* 2018, 24, 1822–1829. 538  
539
13. Roswall, J.; Olsson, L.M.; Kovatcheva-Datchary, P.; Nilsson, S.; Tremaroli, V.; Simon, M.-C.; Kiilerich, P.; Akrami, R.; Krämer, M.; Uhlén, M.; et al. Developmental trajectory of the healthy human gut microbiota during the first 5 years of life. *Cell Host Microbe* **2021**, 1–12, doi:10.1016/j.chom.2021.02.021. 540  
541  
542
14. Albenburg, L.G.; Wu, G.D. Diet and the Intestinal Microbiome: Associations, Functions, and Implications for Health and Disease. *Gastroenterology* **2014**, 146, 1564–1572, doi:10.1053/j.gastro.2014.01.058.Diet. 543  
544
15. Dreyer, J.L.; Liebl, A.L. Early colonization of the gut microbiome and its relationship with obesity. *Hum. Microbiome J.* 2018, 10, 1–5. 545
16. Motherway, M.O.C.; Brien, F.O.; Driscoll, T.O.; Casey, P.G. Carbohydrate Syntrophy enhances the establishment of *Bifidobacterium breve* UCC2003 in the neonatal gut. *Nat. Sci. Reports* **2018**, 8, 1–10, doi:10.1038/s41598-018-29034-0. 546  
547
17. Bokulich, N.A.; Chung, J.; Battaglia, T.; Henderson, N.; Jay, M.; Li, H.; Lieber, A.D.; Wu, F.; Perez-Perez, G.I.; Chen, Y.; et al. Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Microbiome* **2016**, 8. 548  
549
18. Bar, N.; Korem, T.; Weissbrod, O.; Zeevi, D.; Rothschild, D.; Leviatan, S.; Kosower, N.; Lotan-pompan, M.; Weinberger, A.; Roy, C.I. Le; et al. A reference map of potential determinants for the human serum metabolome. *Nature* **2020**, 588, doi:10.1038/s41586-020-2896-2. 550  
551  
552
19. Maruvada, P.; Lampe, J.W.; Wishart, D.S.; Barupal, D.; Chester, D.N.; Dodd, D.; Djoumbou-feunang, Y.; Dorrestein, P.C.; Dragsted, L.O.; Draper, J.; et al. Perspective: Dietary Biomarkers of Intake and Exposure — Exploration with Omics Approaches. *Nutrition* **2020**, 11, 200–215. 553  
554  
555
20. de Souza, R.J.; Shanmuganathan, M.; Lamri, A.; Atkinson, S.A.; Becker, A.; Desai, D.; Gupta, M.; Mandhane, J.; Moraes, T.J.; Morrison, K.M.; et al. Maternal Diet and the Serum Metabolome in Pregnancy: Robust Dietary Biomarkers Generalizable to a Multiethnic Birth Cohort. *Curr. Dev. Nutr.* **2020**, 4, 1–12. 556  
557  
558
21. Cuparencu, C.; Praticó, G.; Hemeryck, L.Y.; Sri Harsha, P.S.C.; Noerman, S.; Rombouts, C.; Xi, M.; Vanhaecke, L.; Hanhineva, K.; Brennan, L.; et al. Biomarkers of meat and seafood intake: an extensive literature review. *Genes Nutr.* **2019**, 1–30. 559  
560
22. Sri Harsha, P.S.C.S.; Wahab, R.A.; Garcia-aloy, M.; Madrid-gambin, F.; Estruel-amades, S.; Watzl, B.; Andrés-lacueva, C.; Brennan, L. Biomarkers of legume intake in human intervention and observational studies: a systematic review. *Genes Nutr.* **2018**, 13, 1–16. 561  
562
23. Brouwer-brolsma, E.M.; Brandl, B.; Buso, M.E.C.; Skurk, T.; Manach, C. Food intake biomarkers for green leafy vegetables, bulb vegetables, and stem vegetables: a review. *Genes Nutr.* **2020**. 563  
564
24. Rafiq, T.; Azab, S.M.; Teo, K.K.; Thabane, L.; Anand, S.S.; Morrison, K.M.; de Souza, R.J.; Britz-McKibbin, P. Nutritional Metabolomics and the Classification of Dietary Biomarker Candidates: A Critical Review. *Adv. Nutr.* **2021**, 1–54. 565  
566
25. Wang, Z.; Klipfell, E.; Bennett, B.J.; Koeth, R.; Levison, B.S.; Dugar, B.; Feldstein, A.E.; Britt, E.B.; Fu, X.; Chung, Y.; et al. Gut flora 567

- metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **2011**, *472*, doi:10.1038/nature09922. 568
26. Deehan, E.C.; Yang, C.; Perez-mun, M.E.; Zhang, Z.; Bakal, J.A.; Deehan, E.C.; Yang, C.; Perez-mun, M.E.; Zhang, Z.; Bakal, J.A.; et al. Precision Microbiome Modulation with Discrete Dietary Fiber Structures Directs Short-Chain Fatty Acid Production. *Cell Host Microbe* **2020**, *27*, 1–16, doi:10.1016/j.chom.2020.01.006. 569  
570  
571
27. Weitkunat, K.; Schumann, S.; Nickel, D.; Hornemann, S.; Petzke, K.J.; Schulze, M.B.; Pfeiffer, A.F.H.; Klaus, S. Odd-chain fatty acids as a biomarker for dietary fiber intake : a novel pathway for endogenous production from propionate 1 , 2. *Am. J. Clin. Nutr.* **2017**, *105*, 1544–1551, doi:10.3945/ajcn.117.152702.INTRODUCTION. 572  
573  
574
28. Anand, S.S.; Vasudevan, A.; Gupta, M.; Morrison, K.; Kurpad, A.; Teo, K.K. Rationale and design of South Asian Birth Cohort ( START ): a Canada-India collaborative study. *BMC Public Health* **2013**, *13*. 575  
576
29. Takaro, T.K.; Scott, J.A.; Allen, R.W.; Anand, S.S.; Becker, A.B.; Befus, A.D.; Brauer, M.; Duncan, J.; Lefebvre, D.L.; Lou, W.; et al. The Canadian Healthy Infant Longitudinal Development (CHILD) birth cohort study: Assessment of environmental exposures. *J. Expo. Sci. Environ. Epidemiol.* **2015**, *25*, 580–592, doi:10.1038/jes.2015.7. 577  
578  
579
30. Kelemen, L.E.; Anand, S.S.; Vuksan, V.; Yi, Q.; Teo, K.K.; Devanesen, S.; Yusuf, S.; Investigators, S. Development and evaluation of cultural food frequency questionnaires for South Asians, Chinese, and Europeans in North America. *J. Am. Diet. Assoc.* **2003**, *103*, doi:10.1053/jada.2003.50578. 580  
581  
582
31. De Souza, R.J.; Williams, N.C.; Sockalingam, L.; Wahli, G.; Desai, D.; Dehghan, M.; Schulze, K.M.; Gupta, M.; Anand, S.S. Validity and Reproducibility of a Semi-Quantitative Food-Frequency Questionnaire Designed to Measure the Nutrient Intakes of Canadian South Asian Infants at 12 Months of Age. *Can. J. Diet. Pract. Res.* **2020**, *81*, 170–178, doi:10.3148/cjdp-2020-011. 583  
584  
585
32. Subbarao, P.; Anand, S.S.; Becker, A.B.; Befus, A.D.; Brauer, M.; Brook, J.R.; Denburg, J.A.; Hayglass, K.T.; Kobor, M.S.; Kollmann, T.R.; et al. The Canadian Healthy Infant Longitudinal Development (CHILD) study: Examining developmental origins of allergy and asthma. *Thorax* **2015**, *70*, 998–1000, doi:10.1136/thoraxjnl-2015-207246. 586  
587  
588
33. Patterson, R.E.; Kristal, A.R.; Tinker, L.F.; Carter, R.A.; Bolton, M.P.; Agurs-Collins, T. Measurement characteristics of the Women’s Health Initiative food frequency questionnaire. *Ann. Epidemiol.* **1999**, *9*, 178–187, doi:10.1016/S1047-2797(98)00055-6. 589  
590
34. Souza, R.J. De; Zulyniak, M.A.; Desai, D.; Shaikh, M.R.; Campbell, N.C.; Lefebvre, D.L.; Gupta, M.; Wilson, J.; Wahli, G.; Atkinson, S.A.; et al. Harmonization of Food-Frequency Questionnaires and Dietary Pattern Analysis in 4 Ethnically Diverse Birth Cohorts. *Nutr. Epidemiol.* **2016**, 2343–2350, doi:10.3945/jn.116.236729.folate. 591  
592  
593
35. Mozaffarian, D. Fish, mercury, selenium and cardiovascular risk: Current evidence and unanswered questions. *Int. J. Environ. Res. Public Health* **2009**, *6*, 1894–1916, doi:10.3390/ijerph6061894. 594  
595
36. Mozaffarian, D. Dairy Foods, Obesity, and Metabolic Health: The Role of the Food Matrix Compared with Single Nutrients. *Adv. Nutr.* **2019**, *10*, 917S–923S, doi:10.1093/advances/nmz053. 596  
597
37. Aune, D.; Giovannucci, E.; Boffetta, P.; Fadnes, L.T.; Keum, N.N.; Norat, T.; Greenwood, D.C.; Riboli, E.; Vatten, L.J.; Tonstad, S. Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality-A systematic review and dose-response meta-analysis of prospective studies. *Int. J. Epidemiol.* **2017**, *46*, 1029–1056, doi:10.1093/ije/dyw319. 598  
599  
600
38. McRae, M.P. Health Benefits of Dietary Whole Grains: An Umbrella Review of Meta-analyses. *J. Chiropr. Med.* **2017**, *16*, 10–18, doi:10.1016/j.jcm.2016.08.008. 601  
602
39. De Souza, R.G.M.; Schincaglia, R.M.; Pimente, G.D.; Mota, J.F. Nuts and human health outcomes: A systematic review. *Nutrients* **2017**, *9*, doi:10.3390/nu9121311. 603  
604
40. Richi, E.B.; Baumer, B.; Conrad, B.; Darioli, R.; Schmid, A.; Keller, U. Health risks associated with meat consumption: A review of epidemiological studies. *Int. J. Vitam. Nutr. Res.* **2015**, *85*, 70–78, doi:10.1024/0300-9831/a000224. 605  
606
41. Borch, D.; Juul-Hindsgaul, N.; Veller, M.; Astrup, A.; Jaskolowski, J.; Raben, A. Potatoes and risk of obesity, type 2 diabetes, and cardiovascular disease in apparently healthy adults: A systematic review of clinical intervention and observational studies. *Am. J. Clin. Nutr.* **2016**, *104*, 489–498, doi:10.3945/ajcn.116.132332. 607  
608  
609

- 
42. Malik, V.S.; Hu, F.B. Sugar-sweetened beverages and cardiometabolic health: An update of the evidence. *Nutrients* **2019**, *11*, 610  
doi:10.3390/nu11081840. 611
43. Al-Ibrahim, A.A.; Jackson, R.T. Healthy eating index versus alternate healthy index in relation to diabetes status and health markers in 612  
U.S. adults: NHANES 2007-2010. *Nutr. J.* **2019**, *18*, 1–18, doi:10.1186/s12937-019-0450-6. 613
44. Cowin, I.; Emmett, P.; Team, A. Diet in a group of 18-month-old children in South West England, and comparison with the results of a 614  
national survey. *J Hum Nutr Diet* **2007**, *67*, 14–21. 615
45. Wright, J.; Small, N.; Raynor, P.; Tuffnell, D.; Bhopal, R.; Cameron, N.; Fairley, L.; A Lawlor, D.; Parslow, R.; Petherick, E.S.; et al. 616  
Cohort profile: The born in bradford multi-ethnic family cohort study. *Int. J. Epidemiol.* **2013**, *42*, 978–991, doi:10.1093/ije/dys112. 617
46. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **2011**, *7*. 618
47. Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Holmes, S.P. DADA2: High resolution sample inference from 619  
Illumina amplicon data. *Nat Methods* **2016**, *13*, 1–7, doi:10.1007/978-1-4614-9209-2\_118-1. 620
48. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA ribosomal RNA gene 621  
database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2013**, *41*, 590–596, doi:10.1093/nar/gks1219. 622
49. Hurlbert, S.H. The Nonconcept of Species Diversity : A Critique and Alternative Parameters. *Ecology* **1971**, *52*, 577–586. 623
50. Hill, M.O. Diversity and Evenness : A Unifying Notation and Its Consequences. *Ecology* **1973**, *54*, 427–432. 624
51. Oksanen, A.J.; Blanchet, F.G.; Friendly, M.; Kindt, R.; Legendre, P.; Mcglinn, D.; Minchin, P.R.; O’Hara, R.B.; Simpson, G.L.; Solymos, 625  
P.; et al. vegan: Community Ecology Package. R package version 2.5-6. **2019**. 626
52. Gloor, G.B.; Macklaim, J.M.; Pawlowsky-glahn, V.; Egozcue, J.J. Microbiome Datasets Are Compositional : And This Is Not Optional. 627  
*Front. Microbiol.* **2017**, *8*, 1–6, doi:10.3389/fmicb.2017.02224. 628
53. Shanmuganathan, M.; Kroezen, Z.; Gill, B.; Azab, S.; de Souza, R.J.; Teo, K.K.; Atkinson, S.; Subbarao, P.; Desai, D.; Anand, S.S.; et 629  
al. The maternal serum metabolome by multisegment injection-capillary electrophoresis-mass spectrometry: a high-throughput platform 630  
and standardized data workflow for large-scale epidemiological studies. *Nat. Protoc.* **2021**, 1–29, doi:10.1038/s41596-020-00475-0. 631
54. Kuehnbaum, N.L.; Kormendi, A.; Britz-Mckibbin, P. Multisegment injection-capillary electrophoresis-mass spectrometry: A high- 632  
throughput platform for metabolomics with high data fidelity. *Anal. Chem.* **2013**, *85*, 10664–10669, doi:10.1021/ac403171u. 633
55. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 634  
**2014**, *15*, 1–21, doi:10.1186/s13059-014-0550-8. 635
56. Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D.S.; Xia, J. MetaboAnalyst 4 . 0 : towards more transparent and 636  
integrative metabolomics analysis. **2018**, *46*, 486–494, doi:10.1093/nar/gky310. 637
57. Garcia-Aloy, M.; Ulaszewska, M.; Franceschi, P.; Estruel-Amades, S.; Weinert, C.H.; Tor-Roca, A.; Urpi-Sarda, M.; Mattivi, F.; Andres- 638  
Lacueva, C. Discovery of Intake Biomarkers of Lentils, Chickpeas, and White Beans by Untargeted LC–MS Metabolomics in Serum and 639  
Urine. *Mol. Nutr. Food Res.* **2020**, *64*, doi:10.1002/mnfr.201901137. 640
58. Garcia-aloy, M.; Hulshof, P.J.M.; Estruel-amades, S.; Osté, M.C.J.; Lankinen, M.; Geleijnse, J.M.; Goede, J. De; Ulaszewska, M.; Mattivi, 641  
F.; Bakker, S.J.L.; et al. Biomarkers of food intake for nuts and vegetable oils : an extensive literature search. *Genes Nutr.* **2019**, *14*, 1– 642  
21. 643
59. Oliphant, K.; Allen-Vercoe, E. Macronutrient metabolism by the human gut microbiome: Major fermentation by-products and their 644  
impact on host health. *Microbiome* **2019**, *7*, doi:10.1186/s40168-019-0704-8. 645
60. Rubio-del-Campo, A.; Alcántara, C.; Collado, M.C.; Rodríguez-Díaz, J.; Yebra, M.J. Human milk and mucosa-associated disaccharides 646  
impact on cultured infant fecal microbiota. *Sci. Rep.* **2020**, *10*, 1–12, doi:10.1038/s41598-020-68718-4. 647
61. Keller, B.O.; Wu, B.T.F.; Li, S.S.J.; Monga, V.; Innis, S.M. Hypaphorine Is Present in Human Milk in Association with Consumption of 648  
Legumes. *J. Agric. Food Chem.* **2013**, doi:10.1021/jf401758f. 649
62. Zarei, I.; Baxter, B.A.; Oppel, R.C.; Borresen, E.C.; Brown, R.J.; Ryan, E.P. Plasma and urine metabolite profiles impacted by increased 650  
dietary navy bean intake in colorectal cancer survivors: A randomized-controlled trial. *Cancer Prev. Res.* **2021**, *14*, 497–508, 651

- doi:10.1158/1940-6207.CAPR-20-0270. 652
63. El-Magd, M.A.; Abdo, W.S.; El-Maddaway, M.; Nasr, N.M.; Gaber, R.A.; El-Shetry, E.S.; Saleh, A.A.; Alzahrani, F.A.A.; Abdelhady, D.H. High doses of S-methylcysteine cause hypoxia-induced cardiomyocyte apoptosis accompanied by engulfment of mitochondria by nucleus. *Biomed. Pharmacother.* **2017**, *94*, 589–597, doi:10.1016/j.biopha.2017.07.100. 653  
654  
655
64. Perera, T.; Young, M.R.; Zhang, Z.; Murphy, G.; Colburn, N.H.; Lanza, E.; Hartman, T.J.; Cross, A.J.; Bobe, G. Identification and monitoring of metabolite markers of dry bean consumption in parallel human and mouse studies. *Mol. Nutr. Food Res.* **2015**, *59*, 795–806, doi:10.1002/mnfr.201400847. 656  
657  
658
65. Mohi El-Din, M.M.; Mostafa, A.M.; Abd-Elkader, A. Experimental studies on the effect of (Lambda-Cyhalothrin) insecticide on lungs and the ameliorating effect of plant extracts (Ginseng (Panax Ginseng) and garlic (Allium sativum L.) on asthma development in albino rats. *BMC Res. Notes* **2014**, *7*, 1–10, doi:10.1186/1756-0500-7-243. 659  
660  
661
66. Lindqvist, H.M.; Rådjursöga, M.; Malmodin, D.; Winkvist, A.; Ellegård, L. Serum metabolite profiles of habitual diet: evaluation by 1H-nuclear magnetic resonance analysis. *Am. J. Clin. Nutr.* **2019**, *110*, 53–62, doi:10.1093/ajcn/nqz032. 662  
663
67. Schmidt, J.A.; Rinaldi, S.; Scalbert, A.; Ferrari, P.; Achaintre, D.; Gunter, M.J.; Appleby, P.N.; Key, T.J.; Travis, R.C. Plasma concentrations and intakes of amino acids in male meat-eaters, fish-eaters, vegetarians and vegans: A cross-sectional analysis in the EPIC-Oxford cohort. *Eur. J. Clin. Nutr.* **2016**, *70*, 306–312, doi:10.1038/ejcn.2015.144. 664  
665  
666
68. Hou, Y.; He, W.; Hu, S.; Wu, G. Composition of polyamines and amino acids in plant-source foods for human consumption. *Amino Acids* **2019**, *51*, 1153–1165, doi:10.1007/s00726-019-02751-0. 667  
668
69. Joshi, J.; Renaud, J.B.; Sumarah, M.W.; Marsolais, F. Deciphering S-methylcysteine biosynthesis in common bean by isotopic tracking with mass spectrometry. *Plant J.* **2019**, *100*, 176–186, doi:10.1111/tj.14438. 669  
670
70. Wang, W.; Wu, Z.; Dai, Z.; Yang, Y.; Wang, J.; Wu, G. Glycine metabolism in animals and humans: Implications for nutrition and health. *Amino Acids* **2013**, *45*, 463–477, doi:10.1007/s00726-013-1493-1. 671  
672
71. Ulaszewska, M.; Vázquez-manjarrez, N.; Garcia-aloy, M.; Llorach, R.; Mattivi, F.; Dragsted, L.O.; Praticò, G.; Manach, C. Food intake biomarkers for apple, pear, and stone fruit. *Genes Nutr.* **2018**, *13*, 1–16. 673  
674
72. Wellington, N.; Shanmuganathan, M.; De Souza, R.J.; Zulyniak, M.A.; Azab, S.; Bloomfield, J.; Mell, A.; Ly, R.; Desai, D.; Anand, S.S.; et al. Metabolic trajectories following contrasting prudent and western diets from food provisions: Identifying robust biomarkers of short-term changes in habitual diet. *Nutrients* **2019**, *11*, doi:10.3390/nu11102407. 675  
676  
677
73. Kochlik, B.; Gerbracht, C.; Grune, T.; Weber, D. The Influence of Dietary Habits and Meat Consumption on Plasma 3-Methylhistidine—A Potential Marker for Muscle Protein Turnover. *Mol. Nutr. Food Res.* **2018**, *62*, 1–9, doi:10.1002/mnfr.201701062. 678  
679
74. Lee, R.; West, D.; Phillips, S.M.; Britz-McKibbin, P. Differential metabolomics for quantitative assessment of oxidative stress with strenuous exercise and nutritional intervention: Thiol-specific regulation of cellular metabolism with N-acetyl-L-cysteine pretreatment. *Anal. Chem.* **2010**, *82*, 2959–2968, doi:10.1021/ac9029746. 680  
681  
682
75. Dabek, A.; Wojtala, M.; Pirola, L.; Balcerczyk, A. Metabolomics by Ketone Bodies . Implications of the Pathological States. *Nutrients* **2020**, *12*. 683  
684
76. Fikri, A.M.; Smyth, R.; Kumar, V.; Al-Abadla, Z.; Abusnana, S.; Munday, M.R. Pre-diagnostic biomarkers of type 2 diabetes identified in the UAE's obese national population using targeted metabolomics. *Sci. Rep.* **2020**, *10*, 1–10, doi:10.1038/s41598-020-73384-7. 685  
686
77. Goldberg, E.L.; Shchukina, I.; Asher, J.L.; Sidorov, S.; Artyomov, M.N.; Dixit, V.D. Ketogenesis activates metabolically protective  $\gamma\delta$  T cells in visceral adipose tissue. *Nat. Metab.* **2020**, *2*, 50–61, doi:10.1038/s42255-019-0160-6. 687  
688
78. Kirchberg, F.F.; Harder, U.; Weber, M.; Grote, V.; Demmelmaier, H.; Peissner, W.; Rzehak, P.; Xhonneux, A.; Carlier, C.; Ferre, N.; et al. Dietary protein intake affects amino acid and acylcarnitine metabolism in infants aged 6 months. *J. Clin. Endocrinol. Metab.* **2015**, *100*, 149–158, doi:10.1210/jc.2014-3157. 689  
690  
691
79. He, X.; Parenti, M.; Grip, T.; Domellöf, M.; Lönnerdal, B.; Hernell, O.; Timby, N.; M. Slupsky, C. Metabolic phenotype of breast-fed infants, and infants fed standard formula or bovine MFGM supplemented formula: a randomized controlled trial. *Sci. Rep.* **2019**, *9*, 1–13, 692  
693

- 
- doi:10.1038/s41598-018-36292-5. 694
80. Shunova, A.; Böckmann, K.A.; Minarski, M.; Franz, A.R.; Wiechers, C.; Poets, C.F.; Bernhard, W. Choline content of term and preterm infant formulae compared to expressed breast milk—how do we justify the discrepancies? *Nutrients* **2020**, *12*, 1–12, doi:10.3390/nu12123815. 695  
696  
697
81. Henrick, B.M.; Rodriguez, L.; Lakshmikanth, T.; Pou, C.; Henckel, E.; Arzoomand, A.; Olin, A.; Wang, J.; Mikes, J.; Tan, Z.; et al. Bifidobacteria-mediated immune system imprinting early in life. *Cell* **2021**, *184*, 3884-3898.e11, doi:10.1016/j.cell.2021.05.030. 698  
699
82. Azab, S.M.; De Souza, R.J.; Teo, K.K.; Anand, S.S.; Williams, N.C.; Holzschuher, J.; McGlory, C.; Philips, S.M.; Britz-McKibbin, P. Serum non-esterified fatty acids have utility as dietary biomarkers of fat intake from fish, fish oil, and dairy in women. *J. Lipid Res.* **2020**, *61*, 933–944, doi:10.1194/JLR.D120000630. 700  
701  
702  
703