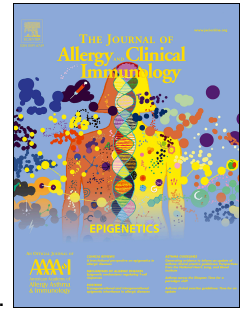


Accepted Manuscript

Integrative Analysis of the Intestinal Metabolome of Childhood Asthma

Kathleen A. Lee-Sarwar, MD, MS, Rachel S. Kelly, PhD, Jessica Lasky-Su, ScD, Robert S. Zeiger, MD, PhD, George T. O'Connor, MD, MS, Megan T. Sandel, MD, Leonard B. Bacharier, MD, Avraham Beigelman, MD, MSCI, Nancy Laranjo, Diane R. Gold, MD, MPH, Scott T. Weiss, MD, MS, Augusto A. Litonjua, MD, MPH



PII: S0091-6749(19)30406-3

DOI: <https://doi.org/10.1016/j.jaci.2019.02.032>

Reference: YMAI 13936

To appear in: *Journal of Allergy and Clinical Immunology*

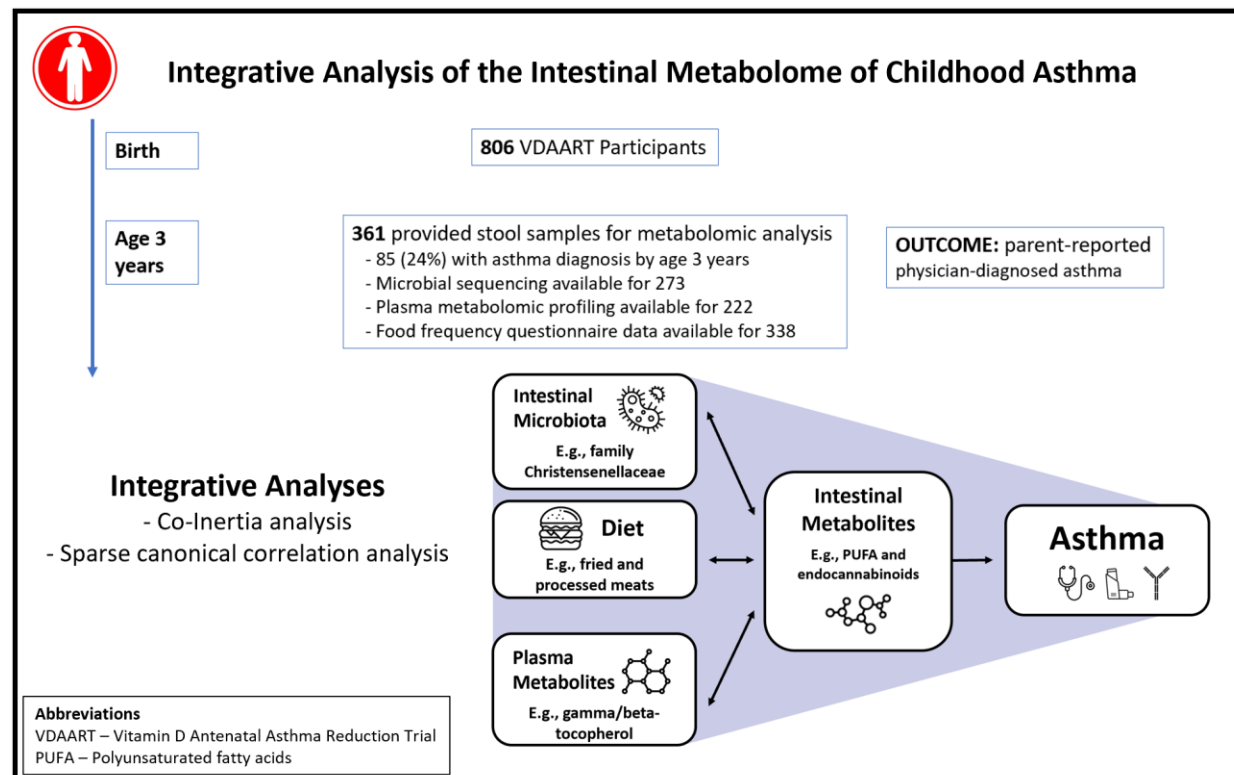
Received Date: 23 October 2018

Revised Date: 14 February 2019

Accepted Date: 20 February 2019

Please cite this article as: Lee-Sarwar KA, Kelly RS, Lasky-Su J, Zeiger RS, O'Connor GT, Sandel MT, Bacharier LB, Beigelman A, Laranjo N, Gold DR, Weiss ST, Litonjua AA, Integrative Analysis of the Intestinal Metabolome of Childhood Asthma, *Journal of Allergy and Clinical Immunology* (2019), doi: <https://doi.org/10.1016/j.jaci.2019.02.032>.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Integrative Analysis of the Intestinal Metabolome of Childhood Asthma

Kathleen A. Lee-Sarwar, MD, MS^{1,2}, Rachel S. Kelly, PhD², Jessica Lasky-Su, ScD², Robert S. Zeiger, MD, PhD³, George T. O'Connor MD, MS⁴, Megan T. Sandel, MD⁵, Leonard B. Bacharier, MD⁶, Avraham Beigelman, MD, MSCI⁶, Nancy Laranjo², Diane R. Gold, MD, MPH^{2,7}, Scott T. Weiss, MD, MS², Augusto A. Litonjua, MD, MPH⁸

¹Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

²Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

³Departments of Allergy and Research and Evaluation, Kaiser Permanente Southern California, San Diego and Pasadena, CA, USA

⁴Pulmonary Center and Department of Medicine, Boston University School of Medicine, Boston, MA, USA

⁵Department of Pediatrics, Boston Medical Center, Boston, MA, USA

⁶Division of Pediatric Allergy, Immunology and Pulmonary Medicine, Department of Pediatrics, Washington University School of Medicine, St Louis, MO and St Louis Children's Hospital, St Louis, MO, USA

⁷Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA

⁸Division of Pediatric Pulmonary Medicine, Golisano Children's Hospital at Strong, University of

Rochester Medical Center, Rochester, NY, USA

Corresponding Authors:

Kathleen A. Lee-Sarwar, MD, MS and Augusto A. Litonjua, MD, MPH

Address:

Brigham and Women's Hospital

Channing Division of Network Medicine

181 Longwood Ave

Boston MA 02115

Telephone: (617) 525-0997

Email: klee-sarwar@partners.org or augusto_litonjua@urmc.rochester.edu

Funding: VDAART was funded by U01HL091528 from the National Heart, Lung, and Blood

Institute. Additional funding came from NIH grants R01HL108818, R01HL123915,

R01HL141826-01, 5T32AI007306-30 and ECHO grant OD023268.

Trial Registration: ClinicalTrials.gov NCT00920621

Conflict of Interest Statement: AAL has received author royalties from UpToDate, Inc. and

consultant fees from AstraZeneca, LP. LBB participates on the Data Safety Monitoring Board of

DBV Technologies. AB holds stock from DBV Technologies. RSZ is a consultant for AstraZeneca, DBV Technologies, Genentech, Inc., GlaxoSmithKline, Novartis, Patara Pharma, Regeneron, TEVA Pharmaceuticals, and Theravance Biopharma, and has received research support from Aerocrine, AstraZeneca, Genentech, Inc., GlaxoSmithKline, NHLBI, MedImmune, and Merck. JL-S is a consultant to Metabolon Inc. GTO is a co-investigator on a grant from Janssen Pharmaceuticals to Boston University that funds a study of the pathogenesis of chronic obstructive pulmonary disease. KL-S, RSK, MS, NL, DRG and STW have nothing to disclose.

Capsule Summary:

Intestinal metabolites, including immune-modulating lipids, are associated with childhood asthma. Specific bacterial taxa, plasma metabolites and dietary factors, including lack of breastfeeding and meat intake, correlate with asthma-associated metabolites.

Key Messages:

- Intestinal metabolites are associated with childhood asthma.
- The intestinal metabolome and microbiome correlate strongly.
- Associations of lack of breastfeeding and meat-rich diet with asthma may be mediated by the intestinal microbiome via intestinal metabolites.

62 **Key Words:** asthma; microbiome; Christensenellaceae; metabolome; vitamin E; nutrition; diet;
63 breastfeeding

64

65 **Abbreviations:**

66 CI – confidence interval

67 DNA – deoxynucleic acid

68 FFQ – food frequency questionnaire

69 IgE – Immunoglobulin E

70 IVF – *in vitro* fertilization

71 PCR – polymerase chain reaction

72 rRNA – ribosomal ribonucleic acid

73 SD – Standard deviation

74 VDAART – Vitamin D Antenatal Asthma Reduction Trial

75 WGCNA – Weighted gene correlation network analysis

ABSTRACT**Background:**

The intestinal metabolome reflects biological consequences of diverse exposures and may provide insight into asthma pathophysiology.

Objective:

To perform an untargeted integrative analysis of the intestinal metabolome of childhood asthma in this ancillary study of the Vitamin D Antenatal Asthma Reduction Trial (VDAART).

Methods:

Metabolomic profiling was performed by mass spectrometry on fecal samples collected from 361 3-year-old subjects. Adjusted logistic regression analyses identified metabolites and modules of highly correlated metabolites associated with asthma diagnosis by age 3 years. Sparse canonical correlation analysis identified associations relevant to asthma between the intestinal metabolome and other “omics”: intestinal microbiome as measured by 16S rRNA sequencing, plasma metabolome as measured by mass spectrometry, and diet as measured by food frequency questionnaires.

Results:

Several intestinal metabolites were associated with asthma at age 3 years, including inverse associations between asthma and polyunsaturated fatty acids (adjusted logistic regression beta

= -6.3, 95% CI -11.3, -1.4, $p = 0.01$) and other lipids. Asthma-associated intestinal metabolites were significant mediators of the inverse relationship between exclusive breastfeeding for the first 4 months of life and asthma (p for indirect association = 0.04), and the positive association between a diet rich in meats and asthma ($p = 0.03$). Specific intestinal bacterial taxa, including family Christensenellaceae, and plasma metabolites, including gamma-tocopherol/beta-tocopherol, were positively associated with asthma and with asthma-associated intestinal metabolites.

Conclusion:

Integrative analyses revealed significant interrelationships between the intestinal metabolome and the intestinal microbiome, plasma metabolome, and diet in association with childhood asthma. These findings require replication in future studies.

INTRODUCTION

Asthma is the most common chronic disease in children(1), and each year approximately 400,000 people die from asthma worldwide(2). Asthma is a complex disease, with dietary, genetic, microbial, and environmental factors all contributing to risk via mechanisms that remain incompletely understood(3). These diverse exposures are reflected by the intestinal metabolome(4), which provides a functional readout of their consequences(5). Intestinal metabolites are small molecules that may be derived from the host, microbiota, or exogenous sources including the diet. Metabolomics analysis has yielded novel insights into the molecular mechanisms of several human diseases(6). Integrative analyses of the human intestinal metabolome with the diet, the intestinal microbiome, and the circulating metabolome reveals intricate relationships, providing insights into the basic physiology of clinically relevant metabolites and their contributions to pathogenic changes(7). For example, in the context of cardiovascular disease, the metabolite trimethylamine-N-oxide is a product of intestinal microbial metabolism of dietary phosphatidylcholine that enters systemic circulation and has been linked to development of atherosclerosis(8). There may be similar undiscovered mechanisms that contribute to the pathophysiology of childhood asthma.

We sought to gain insight into asthma pathophysiology by performing an untargeted analysis of the intestinal metabolome of childhood asthma in this ancillary study of the Vitamin D Antenatal Asthma Reduction Trial (VDAART), hypothesizing that asthma is associated with perturbations to the intestinal metabolome. In addition to characterizing the asthma-associated intestinal metabolome at age 3 years, we conducted integrative analyses to investigate

relationships between the intestinal metabolome and other asthma-related “omics” including the intestinal microbiome, plasma metabolome and diet.

METHODS

All methods are described in detail in the **Supplemental Methods**.

Study Design and Clinical Outcome Ascertainment

Subjects were offspring of participants in VDAART, a multi-site randomized, double-blind, placebo-controlled trial of Vitamin D supplementation during pregnancy for prevention of asthma and other allergic disease in offspring conducted in the United States (NCT00920621)(9). The study protocol was approved by the institutional review boards at each participating institution and at Brigham and Women’s Hospital. All participants provided written informed consent. The asthma outcome was based on parental report of physician diagnosis of asthma in the child’s first 3 years of life. A flow diagram is displayed in **Figure 1**.

Metabolomic Profiling

VDAART participants followed after birth (n = 806) were asked to provide blood and stool samples at age 3 years. Stool samples were provided by 361 (45%) and blood samples by 411 (51%) for metabolomics analysis. Stool was not collected if the child had used antibiotics in the past 7 days. For both stool and plasma samples, metabolomic profiling was performed at Metabolon, Inc. (Research Triangle Park, NC) as previously described(10). Additionally, 245 subjects, 156 of whom had available stool metabolomics, were selected for a prior analysis of plasma metabolomics utilizing an additional quantitative lipids platform at Metabolon, Inc(11). These were included in limited correlation analyses of intestinal and plasma lipids.

Bacterial Microbiome Profiling

DNA extraction and sequencing of the bacterial 16S V4 hyper-variable region using the Illumina MiSeq platform (San Diego, CA) were performed at Partners Healthcare Personalized Medicine (Boston, MA). Additional processing of microbiome data was performed using Qiime(12) and Phyloseq package for R(13). As it was recently shown that quantitative microbiome profiling is preferable to relative abundance profiling in co-occurrence analyses and in seeking disease associations(14), quantitative PCR using universal 16S rRNA primers was performed at Partners Healthcare Personalized Medicine to estimate bacterial biomass concentration per stool sample for use in estimating quantitative species-level taxa abundances.

Dietary Ascertainment

Child diet was evaluated at age 3 years, when parents completed a modified version of a semi-quantitative 87-item food frequency questionnaire (FFQ) that was previously validated in preschool-age children(15). FFQ responses were used to estimate total daily calorie intake as detailed in the **Supplemental Methods**.

Statistics

Details of statistical analyses are provided in the **Supplemental Methods**. Statistical analyses were conducted using R version 3.5.0 (R Foundation for Statistical Computing). Modules of highly correlated and likely functionally related intestinal metabolites were identified using the weighted gene correlation network analysis (WGCNA) R package (version 1.61)(16). Eigenvalues summarizing relative abundances of metabolites of each module for each subject were used in subsequent analyses. Multivariable logistic regression was used to determine the associations of metabolites and metabolite module eigenvalues with asthma. Covariates in adjusted analyses were selected on the basis of significant ($p < 0.05$) bivariate associations between these variables and asthma (**Table 1**) and included sex, race/ethnicity, study center, maternal education and gestational age, which has reported associations both with early microbiome development(17,18) and asthma(19,20). Total IgE, allergic sensitization and eczema were associated with asthma, but were considered to be characteristic of the outcome rather than confounders and as such were not included as covariates.

There are currently no consensus standards for multiple testing correction in metabolomics; methods such as the Bonferroni correction and even more liberal corrections are considered

too stringent for metabolomics data due to the high correlation of functionally related metabolites. We employed a false discovery correction using a “number of effective tests” method that accounts for the highly collinear nature of metabolites(21, 22): we used principal components analysis of the metabolites to determine how many components are required to explain 50% of the variance in the data, with the rationale that correlated metabolites will load similarly to the same components. We consider the number of components, in this case 17, as the number of effective tests, and use this number to compute a multiple testing corrected p value (0.05 divided by $17 = 0.0029$).

Multivariable linear regression analyses were performed to determine associations between potential predictors of the intestinal microenvironment (e.g., breastfeeding, antibiotics exposure, etc.) and asthma-associated metabolites and metabolite modules. Pearson correlation was used to analyze associations between plasma metabolites and corresponding intestinal metabolite module eigenvalues. Mediation analyses were performed for two purposes: (1) to estimate the direct association between breastfeeding and asthma, and the indirect association mediated through intestinal metabolites; and (2) to estimate the direct association between dietary variables and asthma, and the indirect association mediated through intestinal metabolites. All tests were 2-sided and the significance level was pre-specified at $p < .05$ except where otherwise specified.

Integrative Analyses

211 Coinertia analysis was used to compare global associations of the plasma metabolome,
212 intestinal metabolome and diet with the intestinal metabolome. Sparse canonical correlation
213 analysis was used to identify intestinal metabolites that are associated with features of other
214 data types (plasma metabolites, intestinal microbes, and dietary items) and with asthma. These
215 methods are described in detail in the **Supplemental Methods**.

216

RESULTS

Subject Characteristics

To assess the intestinal metabolome of childhood asthma, we analyzed stool samples collected from offspring of VDAART participants. Stool samples were provided by 361 3 year-old subjects, including 85 (24%) diagnosed with asthma by age 3 years (**Table 1**). Subjects with asthma differed from those without asthma on several baseline characteristics among the entire VDAART cohort and similar differences were seen among the subset who provided stool samples. Of subjects who provided stool samples, those with asthma vs. no asthma were significantly more likely to be male (62% vs 54%), black (62% vs 39%), born at earlier gestational ages (mean (SD) 37.8 (2.8) vs 39.0 (2.0) weeks), and born to mothers with lower education (21% vs 37% with college degree or higher) and lower-income households (41% vs 30% income below \$30,000). Asthma status was associated with study site (Boston, St. Louis or San Diego) and this was likely due to differences in race/ethnicity and socioeconomic status between sites. Asthma was also associated with atopy, including elevated serum specific IgE (≥ 0.35 kU/l) to at least one common food or environmental allergen (67% vs 42%), diagnosis of eczema (45% vs 22%), and recent steroid use (29% vs 2%).

Early-Life Intestinal Metabolites are Associated with Childhood Asthma

A total of 737 intestinal metabolites were measured via mass spectrometry and analyzed from stool samples collected at age 3 years. Logistic regression analyses adjusted for sex, race/ethnicity, study center, maternal education and gestational age identified 45 metabolites

that were significantly ($p < 0.05$) associated with asthma by age 3 years (**Table E1**). Of note, all but 3 of these 45 metabolites were inversely associated with asthma. Five of the 45 metabolites remained significant after correcting for multiple testing by a number of effective tests method, including a polyunsaturated fatty acid, omega-6 docosapentaenoic acid, and two diacylglycerols (**Table E1**). As steroid medication use could impact the intestinal metabolome and is expected to be higher in subjects with asthma, we added a covariate to the model to account for use of either inhaled or systemic steroids in the three months prior to turning 3 years old (**Table E1**). Per parental report, 44 (12% of subjects) had used inhaled steroids, 15 (4%) had used systemic steroids; and a total of 48 (13% of subjects) had used either inhaled or systemic steroids. Associations between metabolites and asthma were not significantly changed after adjusting for steroid use (**Table E1**).

To query the functional significance of these metabolites, we constructed a network of highly correlated – and likely functionally related – intestinal metabolites. Network construction was performed on all metabolites with no filtering on association with asthma. The analyses identified 54 modules of highly correlated intestinal metabolites (**Table E2**). Three modules of predominantly lipid metabolites ranging in size from 9 to 24 metabolites were inversely associated with asthma based on their eigenvectors, including a module of primarily diacylglycerols (beta = -6.5, 95% CI = -11.8, -1.4, $p = 0.01$), a module of primarily endocannabinoids (beta = -6.5, 95% CI = -11.6, -1.6, $p = 0.01$) and a module of primarily polyunsaturated fatty acids (PUFAs) (beta = -6.3, 95% CI = -11.3, -1.4, $p = 0.01$) (**Figure 2, Table E3**).

From these analyses, we concluded that several intestinal metabolites were associated with childhood asthma, with the majority of associated metabolites decreased in asthma. Modules of highly correlated asthma-associated lipid metabolites included PUFAs, endocannabinoids and diacylglycerols

Breastfeeding Is a Potential Determinant of Asthma-Associated Intestinal Metabolites

To identify determinants of the asthma-associated intestinal metabolome, we examined potential predictors of the early-life intestinal microenvironment (**Table 2**). In bivariate analyses, those with asthma vs. no asthma were more likely to have been exposed to perinatal antibiotics (52% vs 40%), less likely to have been exclusively breastfed for the first 4 months of life (14% vs 37%) and more likely to have attended daycare by age 3 years (62% vs 50%). We identified no association between asthma and pet dog ownership, mode of delivery or having at least one older sibling. In a logistic regression analysis adjusted for all potential predictors of the early-life intestinal microenvironment and for sex, race/ethnicity, maternal education, study site and gestational age, the only independent predictor of asthma was exclusive breastfeeding for the first 4 months of life, which was inversely associated with asthma (odds ratio 0.36, 95% CI 0.18-0.67, $p = 0.002$).

To investigate the possibility that intestinal metabolites mediate the association between breastfeeding and asthma, we looked for associations between asthma-associated metabolites and breastfeeding in linear regression analyses adjusted for potential confounders. Twelve

intestinal metabolites were associated with asthma and with breastfeeding in a directionally consistent manner (**Table E4**). Mediation analyses estimated that individual metabolites accounted for 9.0% to 23.2% of the association of breastfeeding with asthma, and 8 of the 12 metabolites were significant (p for indirect effect < 0.05) mediators of the association of breastfeeding with asthma (**Table E4**). Seven of the 12 were diacylglycerols, and accordingly, the diacylglycerol metabolite module from network analysis was inversely associated with asthma and positively associated with breastfeeding (beta 0.02, 95% CI 0.01-0.03, $p = 0.004$) in adjusted analyses. Mediation analysis estimated that 13% of the relationship between breastfeeding and asthma was mediated by the diacylglycerol intestinal metabolite module (p for indirect association = 0.04). In sum, this analysis identified lack of exclusive breastfeeding for the first 4 months of life as a risk factor for asthma that may act in part through intestinal metabolites including diacylglycerols.

The Intestinal Metabolome is Globally Associated with the Intestinal Microbiome, Plasma Metabolome, and Diet

We next sought to determine the extent to which the intestinal metabolome associates with other relevant “omics” including the intestinal microbiome, plasma metabolome and diet. The intestinal bacterial microbiome was profiled using quantitative 16S rRNA DNA sequencing on the same stool samples that underwent metabolomic profiling, plasma metabolomic profiling was performed via mass spectrometry, and diet was assessed via food frequency questionnaire responses. The number of subjects with each data type is shown in **Figure 1**.

Coinertia analysis revealed significant global similarity between the intestinal metabolome and other “omics” ($p < 0.02$ for all pairwise “omics” comparisons). This method generates an RV score, which ranges from 0 to 1 with higher scores indicating greater global similarity between a pair of data sets. Diet and the plasma metabolome had similar associations with the intestinal metabolome (RV scores 0.22 and 0.21, respectively), and the intestinal microbiome had the greatest similarity with the intestinal metabolome (RV score 0.35). These results suggest that the intestinal metabolome engages in the most significant interrelationships with the intestinal microbiome, compared to the diet or plasma metabolome.

Intestinal Bacterial Taxa Correlate with Intestinal Metabolites and with Childhood Asthma

We next sought to identify bacterial taxa that correlate with intestinal metabolites and with asthma using sparse canonical correlation analysis, a method that identifies features of two data types (i.e., intestinal metabolome and microbiome) available for the same set of subjects that are correlated with each other and with a clinical outcome (i.e., asthma). This method generates a canonical variate by assigning non-zero canonical loadings to a limited number of intestinal metabolites and taxa, with the direction and magnitude of each feature’s loading indicating the direction and strength of association with asthma and with other features with non-zero loadings. In this analysis, rather than seeking bacteria that are most strongly associated with asthma, we aim to identify highly or uniquely metabolically active bacteria that, by way of metabolic activity, associate with asthma.

Among stool samples collected at age 3 years, a canonical variate was based on 37 intestinal metabolites and 29 bacterial taxa that were associated with each other and with asthma. Canonical loadings for these metabolites and taxa were used to calculate intestinal metabolite and bacterial taxa canonical scores for each subject. Intestinal metabolite and bacterial taxa scores were positively associated with asthma after adjustment for sex, race/ethnicity, study center, maternal education and gestational age (adjusted logistic regression beta = 0.20 and 0.0002; $p = 0.005$ and 0.047 , respectively), and were correlated with each other (Pearson rho = 0.65, $p < 0.001$). Increased abundance of family Christensenellaceae spp was most strongly associated with asthma and asthma-associated intestinal metabolites (**Figure 3**).

When modules of highly correlated metabolites were analyzed instead of individual metabolites, the canonical variate included the same 29 bacterial taxa with similar loadings (**Table E5**) in association with 8 modules of highly correlated metabolites. These 8 modules included the PUFA and endocannabinoid modules that were previously identified as associated with asthma, and which loaded negatively on the canonical variate, consistent with their inverse association with asthma.

These analyses show that there is a strong association between the intestinal metabolome and microbiome that influences asthma risk at age 3 years. We found evidence that PUFAs, endocannabinoids and Christensenellaceae spp participate in relevant microbiome-metabolome interactions.

The Intestinal and Plasma Metabolomes Associate with Childhood Asthma

We also applied sparse canonical correlation analysis to identify 41 intestinal and 37 plasma metabolites from age 3 years that correlated with each other (Pearson $\rho = 0.61$, $p < 0.001$) and with asthma (adjusted logistic regression beta = 0.25 and 1.65; $p = 0.02$ and 0.003, respectively) (**Figure 4**). Interestingly, gamma-tocopherol/beta-tocopherol was the highest-loading feature among both intestinal (loading = 0.31) and plasma (loading = 0.38) metabolites. However, this form of vitamin E, analyzed as an individual metabolite, was not significantly associated with asthma in adjusted logistic regression analyses either in stool (OR 1.43, 95% CI 0.83-2.56, $p = 0.21$) or plasma (OR 1.05, 95% CI 0.11-9.39, $p = 0.97$).

We recently found that plasma PUFAs were inversely associated with asthma at age 3 years in VDAART participants (23). Given the present findings of inverse associations between intestinal PUFA and other lipid metabolite modules and asthma, we hypothesized that plasma and intestinal lipids may be correlated. However, we found no correlation between plasma PUFAs (total, omega-3 or omega-6), endocannabinoids or diacylglycerols and their corresponding intestinal metabolite module eigenvalues (**Table E6**). These findings suggest that intestinal and plasma lipids may relate to asthma risk through different mechanisms.

From these analyses, we concluded that interrelationships between the intestinal and plasma metabolomes associate with asthma at age 3 years. Vitamin E in the form of gamma-tocopherol/beta-tocopherol may, in concert with co-exposures, have a significant influence on the asthma-associated metabolome.

Meat Intake Correlates with Intestinal Metabolites and with Childhood Asthma

Finally, sparse canonical correlation analysis identified 7 foods in the diet and 41 intestinal metabolites that correlated with each other (Pearson $\rho = 0.40$, $p < 0.001$) and with asthma (adjusted logistic regression $\beta = 0.19$ and 0.50 ; $p = 0.01$ and 0.02 , respectively) (**Figure 5**). Asthma associations were preserved after adding BMI and estimated total daily calorie intake to regression models ($\beta = 1.06$, $p = 0.003$ for foods and $\beta = 0.22$, $p = 0.002$ for metabolites). Foods with the highest positive loadings, indicating the most robust positive associations with asthma and intestinal metabolites, included fried and processed meats such as fried chicken, chicken nuggets, hot dogs and hamburgers (**Figure 5**). When modules of highly correlated metabolites were analyzed instead of individual metabolites, the canonical variate included the same 7 foods with similar loadings (**Table E7**) in association with 8 modules of highly correlated metabolites. These 8 modules included the diacylglycerol module that was previously identified as associated with asthma and which loaded negatively on the canonical variate, consistent with its inverse association with asthma. Mediation analysis estimated that 17% of the relationship between dietary score on these 7 foods and asthma was mediated by intestinal metabolite score (p value for indirect association = 0.04). These analyses highlighted an association between a diet rich in meats, especially fried or processed meats, and asthma, with evidence that this association is mediated in part by intestinal metabolites.

Few Intestinal Metabolites Correlate with the Diet, Intestinal Microbiome, Plasma

Metabolome and Asthma

Two metabolites were inversely associated with asthma and had negative loadings in all three sparse canonical correlation analyses (of associations between the intestinal metabolome and the diet, plasma metabolome, and intestinal microbiome): phenol sulfate and galactosylglycerol. These may be central metabolites in asthma pathophysiology; however, given that only 2 such metabolites were found, in general, the intestinal metabolome is likely perturbed in different ways by asthma-associated plasma, dietary and microbiome changes.

Of note, intestinal gamma-tocopherol/beta-tocopherol had positive loadings in all three sparse canonical correlation analyses, and plasma gamma-tocopherol/beta-tocopherol also had a positive loading; however, this form of vitamin E was not associated with asthma when analyzed as an individual metabolite. These findings suggest that this metabolite may act in concert with co-exposures to influence systems-wide effects on asthma-associated “omics.”

DISCUSSION

In this analysis of a relatively large and diverse sample of 3 year-old children, we characterized the intestinal metabolome of asthma. To our knowledge, this is the first report to integrate dietary, microbiome and plasma metabolomics into an evaluation of the intestinal metabolome during childhood. As the intestinal metabolome can be viewed as the functional readout of a variety of exposures, including the diet, the microbiome, and host physiology, we conducted integrative analyses to show that asthma-relevant interactions occur between the intestinal metabolome and the intestinal microbiome, plasma metabolome and diet (**Figure 6**). Our findings, though biologically plausible, require replication in future studies.

Most intestinal metabolites that were associated with asthma were reduced in subjects with asthma, consistent with prior literature showing early-life depletions of specific microbes, microbe functions and metabolites in children with asthma and/or atopy (24, 25). Network analysis revealed inverse associations between asthma and three groups of intestinal lipids: PUFAs, diacylglycerols and endocannabinoids. The role of PUFAs as precursors to immune-modulating eicosanoids and pro-resolving mediators lends biological plausibility to an association with asthma(26). Indeed, dietary and circulating PUFAs have been associated with asthma, including in the VDAART cohort(23). However, we found no association between plasma PUFA and intestinal PUFA in the present analysis. While members of the human intestinal microbiota are not known to produce PUFAs, some participate in metabolism of PUFAs(27, 28). So, intestinal PUFA likely represents dietary PUFA that is both in excess of the amount absorbed in the small intestine, and that is not metabolized by the microbiota; that is, intestinal PUFA abundance is determined both by dietary intake and microbiota composition. Accordingly, we found that intestinal PUFA correlated with the asthma-associated intestinal microbiome. These findings suggest that intestinal and plasma PUFAs relate to asthma risk through different mechanisms, with the former more reflective of bacterial dysbiosis.

Diacylglycerols are crucial intracellular signaling second messengers involved in functions including immune cell signaling and cell growth(29). Intestinal extracellular diacylglycerols influence colonic epithelial cell growth in the setting of malignancy(30) and may be taken up from the intestinal lumen by host cells in other contexts. We report a novel association

between intestinal diacylglycerols and childhood asthma, and provide evidence that intestinal diacylglycerols mediate associations between dietary factors and asthma. Specifically, fecal diacylglycerols were mediators of the inverse association between breastfeeding and asthma. Breastfeeding has long-term influences on the intestinal microenvironment lasting at least to school age (31). Though there have been several reports of inverse associations between breastfeeding and asthma, this association has been inconsistent(32-35). Some of this inconsistency could be explained by differences in other determinants of the intestinal microenvironment, such as the microbiota and subsequent diet, both of which influence fecal diacylglycerol abundances(36). Consistent with this, we found that intestinal diacylglycerols were associated with an asthma-associated diet rich in fried and processed meats, providing evidence that both breastfeeding and subsequent diet may influence asthma risk by way of their effects on the intestinal metabolome. Diacylglycerols are used as emulsifiers in a wide range of foods, though we cannot determine if the fecal diacylglycerols in our sample are of dietary, microbial or host origin.

Endocannabinoids have both dietary and endogenous sources and are involved with broad systemic processes including regulation of immune and metabolic functions(37). In addition to an inverse association between intestinal endocannabinoids and asthma, we found that intestinal endocannabinoids correlated with the asthma-associated microbiome. The mechanism of these novel potential associations is unclear, and one could speculate involvement of effects on mast cell activation(38), PPAR-alpha activity(39), or eicosanoid mediators(40).

We found that the intestinal microbiome had a strong global association with the intestinal metabolome, in keeping with other reports(5). In particular, bacteria of the family Christensenellaceae were associated with asthma and asthma-associated intestinal metabolites. Christensenellaceae has been most strongly associated with lean (vs obese) weight(41, 42). To our knowledge, Christensenellaceae spp. have not been previously associated with asthma, and this may be in part because prior analyses of the intestinal microbiome of asthma have focused on the microbiome in infancy(24, 25, 43–44) or adulthood(45), in contrast to our study of early childhood. Interestingly, an analysis of TwinsUK data found that the family Christensenellaceae was the most heritable taxon analyzed(41), raising the possibility that it could be a marker or consequence of genetic variants that associate with asthma.

We identified a potential multi-systems role for intestinal gamma-tocopherol/beta-tocopherol, which had positive loadings in all sparse canonical correlation analyses; however, effects are likely indirect, as this form of vitamin E was not associated with asthma when analyzed as an individual metabolite. Vitamin E has anti-oxidant properties and has been inconsistently linked to reduced risk of asthma(46). In contrast and consistent with our findings, the gamma-tocopherol form – which has increased in Western diets over recent decades(47) – has been associated with increased asthma risk or severity(47–49), though some analyses find anti-inflammatory effects(50). The systems effects of dietary vitamin E intake on the microbiome and metabolome are worthy of further investigation and may have relevance to asthma.

480
481 Finally, we found that a diet rich in fried and processed meats was associated with asthma and
482 with asthma-associated intestinal metabolites. Several prior studies have found positive
483 associations between processed or fried meat consumption and asthma(51–53), and one recent
484 systematic review found that hamburger consumption had a particularly strong association
485 with asthma(54). Our findings support a causal pathway whereby intestinal metabolites
486 mediate at least part of the association between fried and processed meat consumption and
487 asthma.

488
489 Some study limitations occurred. The outcome of asthma was based on parent-reported
490 asthma diagnosis by age 3 years. Though preschool asthma is clinically important, this outcome
491 may not accurately capture subjects whose asthma persists later in life(55). We expect this
492 potential misclassification to result in a bias toward the null and reduced power to detect
493 relevant associations. Generalizability may be limited by the VDAART study inclusion criteria,
494 which included parental history of asthma, eczema or allergic rhinitis; and maternal non-
495 smoking status. Intestinal metabolomics were not available for all VDAART subjects, though
496 baseline characteristics were similar between subjects with available intestinal metabolomics
497 and the overall cohort (**Table 1**). Although subjects who had used antibiotics in the week prior
498 to stool sample collection were excluded and we have performed sensitivity analyses to
499 account for steroid use, we cannot rule out residual confounding by medication use. While we
500 adjusted for maternal education, we acknowledge the possibility of residual confounding by
501 other socioeconomic factors. Current metabolomics analysis techniques do not reflect the

entire intestinal chemical repertoire. Some metabolites, including most short chain fatty acids, are too volatile or low-concentration to be measured by the utilized metabolomic profiling method. Dietary analysis by food frequency questionnaire, though widely used, is subject to measurement bias. Causality cannot be inferred from these cross-sectional associations in 3 year-old subjects, and identified associations may not be relevant earlier or later in life. As VDAART participants continue to age, it will be important to follow up our findings with longitudinal analyses.

In summary, we found that intestinal metabolites are associated with asthma at age 3 years. Lack of exclusive breastfeeding for the first 4 months of life was associated with increased risk of asthma and may act via perturbation of the intestinal metabolome. Integrative analyses revealed that the intestinal metabolome is strongly associated with the intestinal microbiome, and with regard to asthma, species of the family Christensenellaceae may play a role in this relationship. A diet rich in fried or processed meats correlates with asthma in an association that may be mediated in part by intestinal metabolites. And finally, vitamin E in the form of gamma-tocopherol/beta-tocopherol may have indirect but influential systems-wide effects on asthma risk. These findings are situated within a prior literature supporting their biological relevance, and are worthy of further exploration to generate novel asthma biomarkers and treatment targets.

REFERENCES

1. Asher I, Pearce N. Global burden of asthma among children. *Int J Tuberc Lung Dis* 2014;18:1269–1278.
2. GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Respir Med* 2017;5:691–706.
3. Beasley R, Semprini A, Mitchell EA. Risk factors for asthma: Is prevention possible? *Lancet* 2015;386:1075–1085.
4. Marcobal A, Kashyap PC, Nelson TA, Aronov PA, Donia MS, Spormann A, et al. A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice. *ISME J* 2013;7:1933–1943.
5. Zierer J, Jackson MA, Kastenmüller G, Mangino M, Long T, Telenti A, et al. The fecal metabolome as a functional readout of the gut microbiome. *Nat Genet* 2018;50(6):790–795.
6. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* 2016;17:451–9.
7. McHardy IH, Goudarzi M, Tong M, Ruegger PM, Schwager E, Weger JR, et al. Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. *Microbiome* 2013;1:17.
8. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 2011;472:57–65.

- 544 9. Litonjua AA, Lange NE, Carey VJ, Brown S, Laranjo N, Harshfield BJ, et al. The Vitamin D
545 Antenatal Asthma Reduction Trial (VDAART): Rationale, design, and methods of a
546 randomized, controlled trial of vitamin D supplementation in pregnancy for the primary
547 prevention of asthma and allergies in children. *Contemp Clin trials* 2014;38:37–50.
- 548 10. Evans AM, Bridgewater BR, Liu Q, Mitchell MW, Robinson RJ, Dai H, et al. Metabolomics:
549 Open access high resolution mass spectrometry improves data quantity and quality as
550 compared to unit mass resolution mass spectrometry in high-throughput profiling
551 metabolomics. *Metabolomics* 2014;4:1–7.
- 552 11. Blighe K, Chawes BL, Kelly RS, Mirzakhani H, McGeachie M, Litonjua AA, et al. Vitamin D
553 prenatal programming of childhood metabolomics profiles at age 3 y. *Am J Clin Nutr*
554 2017;106:1092–1099.
- 555 12. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME
556 allows analysis of high-throughput community sequencing data. *Nat Methods*
557 2010;7:335–6.
- 558 13. McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive Analysis
559 and Graphics of Microbiome Census Data. *PLoS One* 2013;8:e61217.
- 560 14. Vandeputte D, Kathagen G, D’Hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, et al.
561 Quantitative microbiome profiling links gut community variation to microbial load.
562 *Nature* 2017;551:507–511.
- 563 15. Blum RE, Wei EK, Rockett HR, Langeliers JD, Leppert J, Gardner JD, et al. Validation of a
564 food frequency questionnaire in Native American and Caucasian children 1 to 5 years of
565 age. *Matern Child Health J* 1999;3:167–72.

- 566 16. Langfelder P, Horvath S, Fisher R, Zhou X, Kao M, Wong W, *et al.* WGCNA: an R package
567 for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
- 568 17. Forsgren M, Isolauri E, Salminen S, Rautava S. Late preterm birth has direct and indirect
569 effects on infant gut microbiota development during the first six months of life. *Acta*
570 *Paediatr.* 2017;106:1103-1109.
- 571 18. Chernikova DA, Madan JC, Housman ML, Zain-Ul-Abideen M, Lundgren SN, Morrison HG,
572 *et al.* The premature infant gut microbiome during the first 6 weeks of life differs based
573 on gestational maturity at birth. *Pediatr Res.* 2018;84:71-79.
- 574 19. Källén B, Finnström O, Nygren KG, Otterblad Olausson P. Association between preterm
575 birth and intrauterine growth retardation and child asthma. *Eur Respir J.* 2013;41:671-6.
- 576 20. Jaakkola JJ, Ahmed P, Ieromnimon A, Goepfert P, Laiou E, Quansah R, *et al.* Preterm
577 delivery and asthma: a systematic review and meta-analysis. *J Allergy Clin Immunol.*
578 2006;118:823.
- 579 21. Li, M.X., Yeung JM, Cherny SS, Sham PC. Evaluating the effective numbers of independent
580 tests and significant p-value thresholds in commercial genotyping arrays and public
581 imputation reference datasets. *Hum Genet* 2012;131:747-56.
- 582 22. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms
583 in linkage disequilibrium with each other. *Am J Hum Genet* 2004;74:765-9.
- 584 23. Lee-Sarwar KA, Kelly RS, Lasky-Su J, Kachroo P, Zeiger RS, O'Connor GT, *et al.* Dietary and
585 plasma polyunsaturated fatty acids are inversely associated with asthma and atopy in
586 early childhood. *J Allergy Clin Immunol In Pract.* (In press)
- 587 24. Fujimura KE, Sitarik AR, Havstad S, Lin DL, Levan S, Fadrosch D, *et al.* Neonatal gut

microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nat Med* 2016;22:1187–1191.

25. Arrieta M-C, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci Transl Med* 2015;7:307ra152-307ra152.

26. Miles E, Calder P. Can Early Omega-3 Fatty Acid Exposure Reduce Risk of Childhood Allergic Disease? *Nutrients* 2017; 9(7).E784.

27. Moi IM, Leow ATC, Ali MSM, Rahman RNZRA, Salleh AB, Sabri S. Polyunsaturated fatty acids in marine bacteria and strategies to enhance their production. *Appl Microbiol Biotechnol* 2018;102:5811-26.

28. Kishino S, Takeuchi M, Park S-B, Hirata A, Kitamura N, Kunisawa J, et al. Polyunsaturated fatty acid saturation by gut lactic acid bacteria affecting host lipid composition. *Proc Natl Acad Sci* 2013;110:17808–13.

29. Eichmann TO, Lass A. DAG tales: The multiple faces of diacylglycerol – stereochemistry, metabolism, and signaling. *Cell Mol Life Sci* 2015;72(20):3931-52.

30. Friedman E, Isaksson P, Rafter J, Marian B, Winawer S, Newmark H. Fecal diglycerides as selective endogenous mitogens for premalignant and malignant human colonic epithelial cells. *Cancer Res* 1989;49:544–548.

31. Zhong H, Penders J, Shi Z, Ren H, Cai K, Fang C, et al. Impact of early events and lifestyle on the gut microbiota and metabolic phenotypes in young school-age children. *Microbiome* 2019;7:2.

32. Dogaru CM, Nyffenegger D, Pescatore AM, Spycher BD, Kuehni CE. Breastfeeding and

childhood asthma: systematic review and meta-analysis. *Am J Epidemiol* 2014;179:1153–67.

33. Lodge C, Tan D, Lau M, Dai X, Tham R, Lowe A, et al. Breastfeeding and asthma and allergies: a systematic review and meta-analysis. *Acta Paediatr* 2015;104:38–53.

34. Scholtens S, Wijga AH, Brunekreef B, Kerkhof M, Hoekstra MO, Gerritsen J, et al. Breast feeding, parental allergy and asthma in children followed for 8 years. The PIAMA birth cohort study. *Thorax* 2009;64:604–609.

35. Oddy WH, Holt PG, Sly PD, Read AW, Lazndau LI, Stanley FJ, et al. Association between breast feeding and asthma in 6 year old children: findings of a prospective birth cohort study. *Br Med J* 1999;319:815–819.

36. Vulevic J, McCartney AL, Gee JM, Johnson IT, Gibson GR. Microbial species involved in production of 1,2-sn-diacylglycerol and effects of phosphatidylcholine on human fecal microbiota. *Appl Environ Microbiol* 2004;70:5659–66.

37. Iannotti FA, Di Marzo V, Petrosino S. Endocannabinoids and endocannabinoid-related mediators: Targets, metabolism and role in neurological disorders. *Prog Lipid Res* 2016;62:107-28.

38. Roviezzo F, Rossi A, Caiazzo E, Orlando P, Riemma MA, Iacono VM, et al. Palmitoylethanolamide supplementation during sensitization prevents airway allergic symptoms in the mouse. *Front Pharmacol* 2017;8:857.

39. Delayre-Orthez C, Becker J, Auwerx J, Frossard N, Pons F. Suppression of allergen-induced airway inflammation and immune response by the peroxisome proliferator-activated receptor-alpha agonist fenofibrate. *Eur J Pharmacol* 2008;581:177–184.

40. Turcotte C, Chouinard F, Lefebvre JS, Flamand N. Regulation of inflammation by cannabinoids, the endocannabinoids 2-arachidonoyl-glycerol and arachidonoyl-ethanolamide, and their metabolites. *J Leukoc Biol* 2015;97:1049–1070.
41. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. *Cell* 2014;159:789–799.
42. Peters BA, Shapiro JA, Church TR, Miller G, Trinh-Shevrin C, Yuen E, et al. A taxonomic signature of obesity in a large study of American adults. *Sci Rep* 2018;8:9749.
43. Stokholm J, Blaser MJ, Thorsen J, Rasmussen MA, Waage J, Vinding RK, et al. Maturation of the gut microbiome and risk of asthma in childhood. *Nat Commun* 2018;9:141.
44. Stiemsma LT, Arrieta M-C, Dimitriu PA, Cheng J, Thorson L, Lefebvre DL, et al. Shifts in *Lachnospira* and *Clostridium* sp. in the 3-month stool microbiome are associated with preschool age asthma. *Clin Sci* 2016;130:2199–2207.
45. Hevia A, Milani C, López P, Donado CD, Cuervo A, González S, et al. Allergic patients with long-term asthma display low levels of *Bifidobacterium adolescentis*. *PLoS One* 2016;11(2):e0147809.
46. Wu H, Zhang C, Wang Y, Li Y. Does vitamin E prevent asthma or wheeze in children: A systematic review and meta-analysis. *Paediatr Respir Rev* 2018;27:60-8.
47. Cook-Mills JM, Abdala-Valencia H, Hartert T. Two faces of vitamin E in the lung. *Am J Respir Crit Care Med* 2013;188:279–284.
48. Marchese ME, Kumar R, Colangelo LA, Avila PC, Jacobs DR, Gross M, et al. The vitamin E isoforms α -tocopherol and γ -tocopherol have opposite associations with spirometric parameters: The CARDIA study. *Respir Res* 2014;15:31.

49. Cook-Mills JM, Gebretsadik T, Abdala-Valencia H, Green J, Larkin EK, Dupont WD, et al. Interaction of Vitamin E isoforms on asthma and allergic airway disease. *Thorax* 2016;71:954–956.
50. Burbank AJ, Duran CG, Pan Y, Burns P, Jones S, Jiang Q, et al. Gamma tocopherol-enriched supplement reduces sputum eosinophilia and endotoxin-induced sputum neutrophilia in volunteers with asthma. *J Allergy Clin Immunol* 2018;141:1231–1238.e1.
51. Gomes De Luna MDF, Gomes De Luna JR, Fisher GB, De Almeida PC, Chiesa D, Carlos Da Silva MG. Factors associated with asthma in adolescents in the city of Fortaleza, Brazil. *J Asthma* 2015;52:485–491.
52. Cepeda AM, Thawer S, Boyle RJ, Villalba S, Jaller R, Tapias E, et al. Diet and respiratory health in children from 11 Latin American countries: Evidence from ISAAC Phase III. *Lung* 2017;195:683–692.
53. Melo B, Rezende L, Machado P, Gouveia N, Levy R. Associations of ultra-processed food and drink products with asthma and wheezing among Brazilian adolescents. *Pediatr Allergy Immunol* 2018;29:504-11.
54. Wang C, Wang J, Zhang X, Zhang L, Zhang H, Wang L, et al. Is the consumption of fast foods associated with asthma or other allergic diseases? *Respirology* 2018;23:901-13.
55. Bacharier LB, Guilbert TW. Diagnosis and management of early asthma in preschool-aged children. *J Allergy Clin Immunol* 2012;130:287-96.
56. Litonjua AA, Carey VJ, Laranjo N, Harshfield BJ, McElrath TF, O'Connor GT, et al. Effect of prenatal supplementation with vitamin D on asthma or recurrent wheezing in offspring by age 3 years: The VDAART randomized clinical trial. *JAMA* 2016;315:362–70.

FIGURE LEGENDS

Figure 1. Flow diagram of subjects included in study. A portion of this figure has been published previously(56).

Figure 2. Members of the endocannabinoid (A), polyunsaturated fatty acid (B) and diacylglycerol (C) metabolite modules, which were inversely associated with asthma at age 3 years, are displayed by their association with asthma (y axis, adjusted odds ratio) and by module membership (correlation of individual metabolite with module eigenvalue, x axis).

Figure 3. Spearman correlation heatmap of intestinal metabolites (rows) and intestinal bacterial taxa (columns) that are correlated with each other and with asthma at age 3 years in sparse canonical correlation analysis. Canonical loadings are given for each feature after the feature name, with higher loadings indicating greater contribution to the canonical variate and sign indicating direction of association. Red indicates negative correlations and yellow indicates positive correlations. This analysis included 273 subjects with both plasma and stool metabolomics data available.

Figure 4. Spearman correlation heatmap of intestinal metabolites (rows) and plasma metabolites (columns) that are correlated with each other and with asthma at age 3 years in sparse canonical correlation analysis. Canonical loadings are given for each feature after the

feature name, with higher loadings indicating greater contribution to the canonical variate and sign indicating direction of association. Red indicates negative correlations and yellow indicates positive correlations. This analysis included 222 subjects with both plasma and stool metabolomics data available.

Figure 5. Spearman correlation heatmap of foods (rows) and intestinal metabolites (columns) that are correlated with each other and with asthma at age 3 years in sparse canonical correlation analysis. Canonical loadings are given for each feature after the feature name, with higher loadings indicating greater contribution to the canonical variate and sign indicating direction of association. Red indicates negative correlations and yellow indicates positive correlations. This analysis included 338 subjects with food frequency questionnaire and stool metabolomics data available.

Figure 6. Schematic of results of integrative analyses of the asthma-associated intestinal metabolome. Note that a proposed causal pathway is depicted here and is one of several potential causal pathways that could explain the observed associations.

Table 1. Baseline characteristics. Tabulated for subjects who provided stool samples for metabolomics analysis and for the entire VDAART offspring cohort.

	Stool Metabolomics Available				Entire VDAART Cohort			
	All children (n = 361)	Asthma (n = 85)	No Asthma (n = 276)	p value	All children (n = 806)	Asthma (n = 125)	No Asthma (n = 623)	p value
Gender – number (%)				0.24				0.02
Male	203 (56)	53 (62)	150 (54)		421 (52)	79 (63)	318 (51)	
Female	158 (44)	32 (38)	126 (46)		385 (48)	46 (37)	305 (49)	
Race/ethnicity – number (%)				0.001				<0.001
Black, non-Hispanic	161 (45)	53 (62)	108 (39)		317 (39)	73 (58)	229 (37)	
White, non-Hispanic	71 (20)	13 (15)	58 (21)		161 (20)	17 (14)	130 (21)	
Hispanic or Other	129 (36)	19 (22)	110 (40)		328 (41)	35 (28)	264 (42)	
Gestational age (weeks) – mean (SD)	38.7 (2.5)	37.8 (2.8)	39.0 (2.0)	<0.001	39.0 (2.0)	38.3 (2.6)	39.1 (1.8)	<0.001
VDAART treatment group – number (%)				0.87				0.64
4,400 IU/day Vitamin D	182 (50)	44 (52)	138 (50)		405 (50)	60 (48)	316 (51)	
400 IU/day Vitamin D	179 (50)	41 (48)	138 (50)		401 (50)	65 (52)	307 (49)	
Study Center – number (%)				0.006				0.001
Boston	82 (23)	19 (22)	63 (23)		240 (30)	39 (31)	175 (28)	
St. Louis	178 (49)	53 (62)	125 (45)		292 (36)	62 (50)	227 (36)	
San Diego	101 (28)	13 (15)	88 (32)		274 (34)	24 (19)	221 (35)	
Maternal education – number (%)				0.04				0.005
< High school	50 (14)	16 (19)	34 (12)		100 (12)	23 (18)	73 (12)	
High school or Technical School	106 (29)	27 (32)	79 (29)		241 (30)	40 (32)	181 (29)	
Some college	84 (23)	24 (28)	60 (22)		192 (24)	35 (28)	137 (22)	
College graduate or higher	121 (34)	18 (21)	103 (37)		273 (34)	27 (22)	232 (37)	
Household Income – number (%)				0.02				0.004
< \$30,000	118 (33)	35 (41)	83 (30)		236 (29)	47 (38)	176 (28)	
\$30,000-\$49,999	42 (12)	10 (12)	32 (12)		105 (13)	17 (14)	78 (13)	
\$50,000-\$74,999	38 (11)	9 (11)	29 (11)		97 (12)	12 (10)	77 (12)	
\$75,000-\$99,999	29 (8)	3 (4)	26 (9)		79 (10)	5 (4)	68 (11)	
\$100,000-\$149,999	38 (11)	4 (5)	34 (12)		62 (8)	6 (5)	53 (9)	
> \$150,000	16 (4)	0 (0)	16 (6)		31 (4)	0 (0)	29 (5)	
Refused to say or unknown	80 (22)	24 (28)	56 (20)		196 (24)	38 (30)	142 (23)	
Total IgE – mean (SD)	135 (494)	286 (941)	87 (183)	<0.001	123 (434)	273 (882)	96 (271)	0.003
Allergic sensitization – number (%)	145 (48)	49 (67)	96 (42)	<0.001	253 (47)	55 (63)	198 (44)	0.001
Eczema – number (%)	100 (28)	38 (45)	62 (22)	<0.001	172 (23)	43 (36)	117 (19)	<0.001
Recent Steroid Use – number (%)	31 (9)	25 (29)	6 (2)	<0.001	48 (6)	39 (31)	9 (1)	<0.001

p values are for t tests for gestational age and total IgE; and for Chi square tests for all other comparisons. Total IgE was log-transformed before comparison by t test. Recent steroid use pertains to use of inhaled or systemic steroids in the three months prior to the age 3 years study visit, during which stool samples were collected.

Missing data: Asthma status was missing for 58 subjects, none of whom provided stool samples for metabolomics. Gestational age was unavailable for 2 children. Total and serum specific IgE

721 was missing for 272 and 265 subjects, respectively; including 62 and 59 subjects with stool
722 metabolomics. Eczema at age 3 years was unavailable for 55 subjects, none of whom provided
723 stool samples for metabolomics.
724

Table 2. Bivariate analyses of the associations of asthma with potential determinants of the early-life intestinal metabolome. Analyses were performed for subjects who provided stool samples for metabolomics analysis and for the entire VDAART offspring cohort.

	Stool Metabolomics Available				Entire VDAART Cohort			
	All children (n = 361)	Asthma (n = 85)	No Asthma (n = 276)	p value	All children (n = 806)	Asthma (n = 125)	No Asthma (n = 623)	p value
Maternal antibiotics during or child antibiotics after delivery – number (%)	157 (43)	45 (53)	112 (41)	0.06	337 (42)	65 (52)	248 (40)	0.01
Birth by Cesarean section – number (%)	113 (31)	24 (28)	89 (32)	0.57	239 (30)	40 (32)	185 (30)	0.65
Exclusive breastfeeding until age 4 months – number (%)	109 (31)	12 (15)	97 (36)	0.001	247 (33)	16 (14)	218 (37)	<0.001
Day care by age 3 years – number (%)	178 (50)	47 (57)	131 (49)	0.24	378 (53)	73 (62)	290 (50)	0.02
Pet dog during infancy – number (%)	88 (25)	17 (20)	71 (26)	0.38	168 (22)	21 (18)	142 (24)	0.20
At least one living child previously born to mother – number (%)	205 (57)	50 (59)	155 (56)	0.76	435 (54)	67 (54)	344 (55)	0.82

p values are for Chi square tests.

Unavailable/Missing Data: Perinatal antibiotics and mode of delivery were unavailable for 3 subjects. Exclusive breastfeeding until age 4 months was unavailable for 68 subjects, including 13 with available stool metabolomics. Daycare status was unavailable for 99 subjects, including 8 with available stool metabolomics. Household dog status was unavailable for 49 subjects, including 4 with available stool metabolomics.

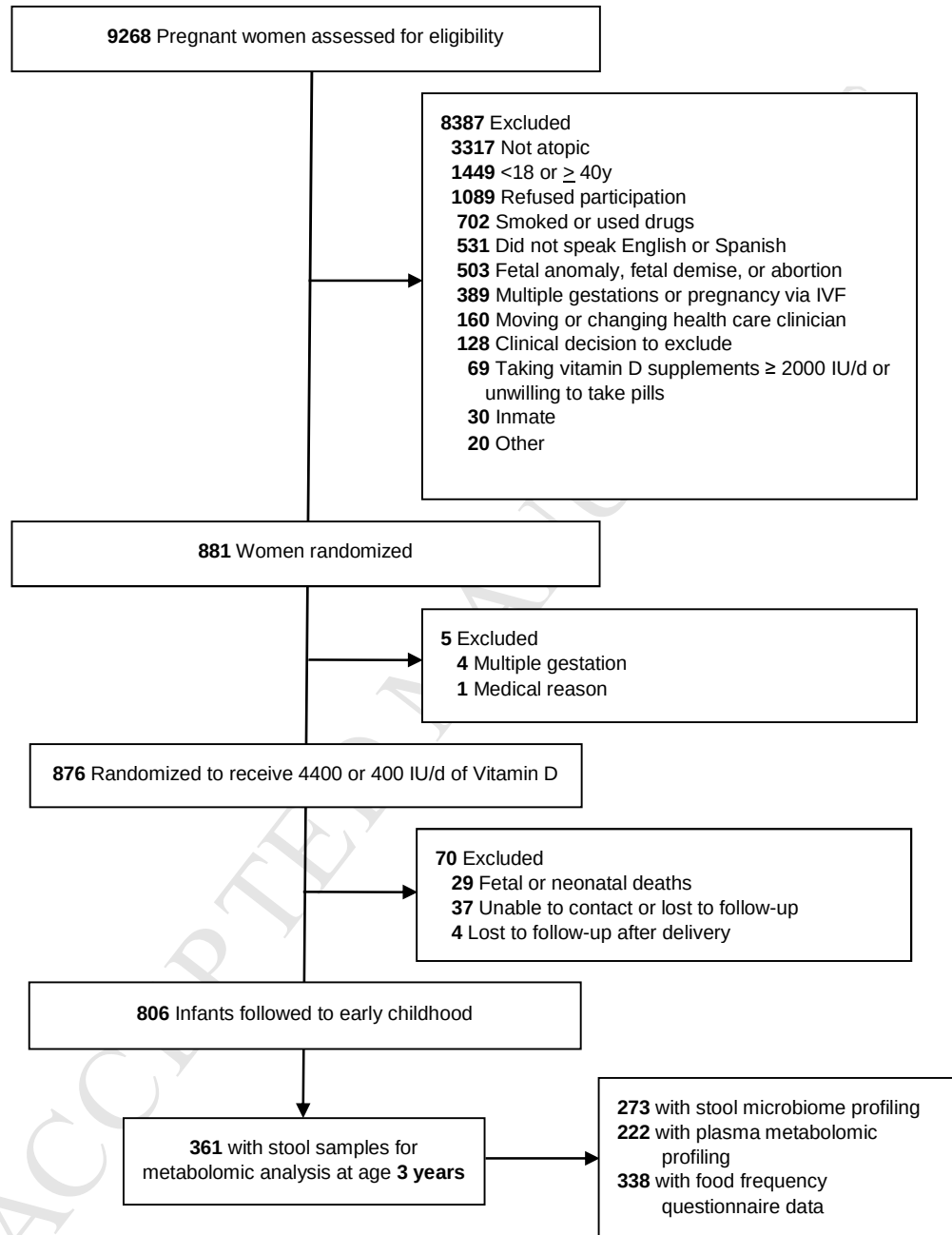
Figure 1.

Figure 2

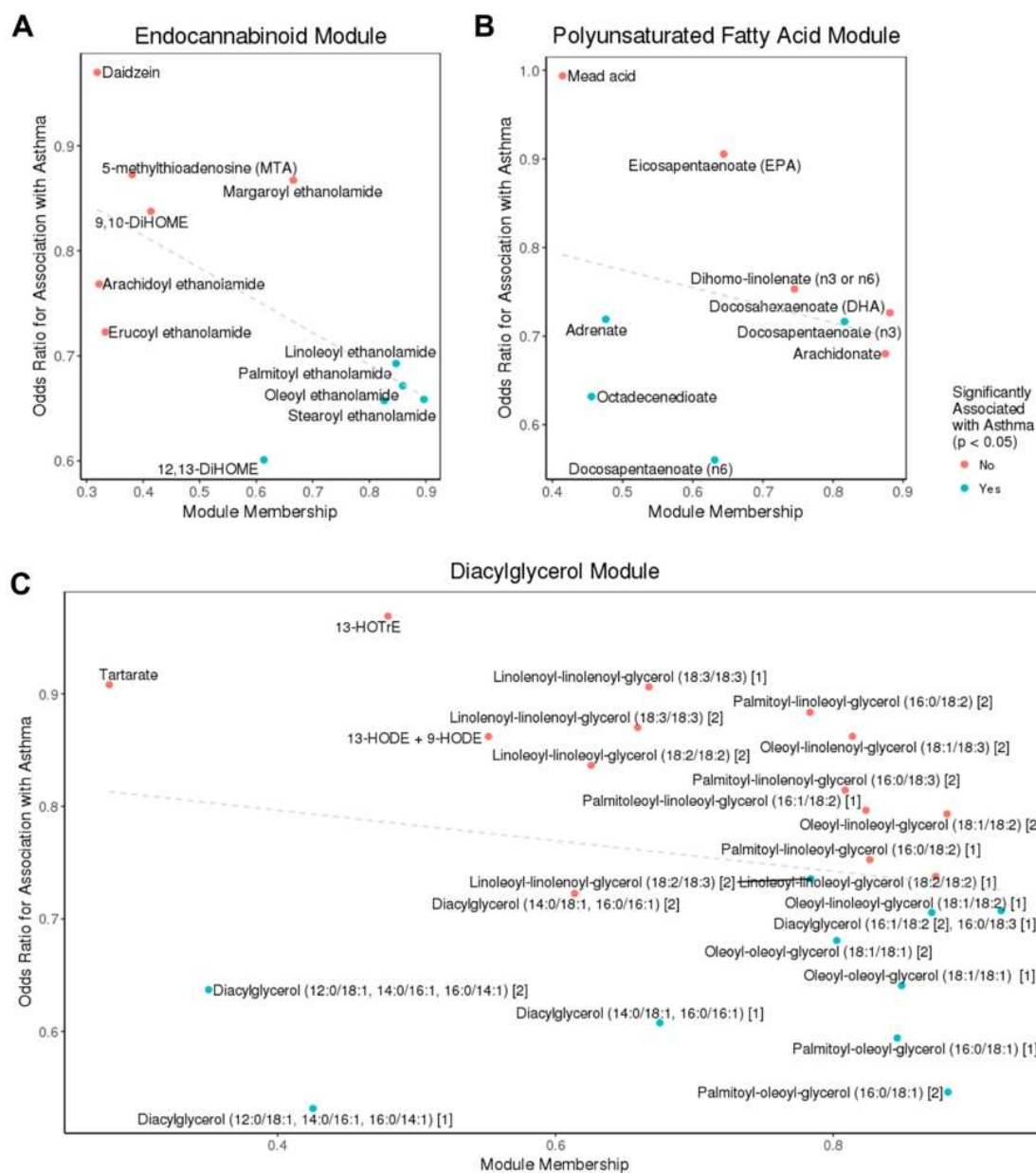


Figure 3

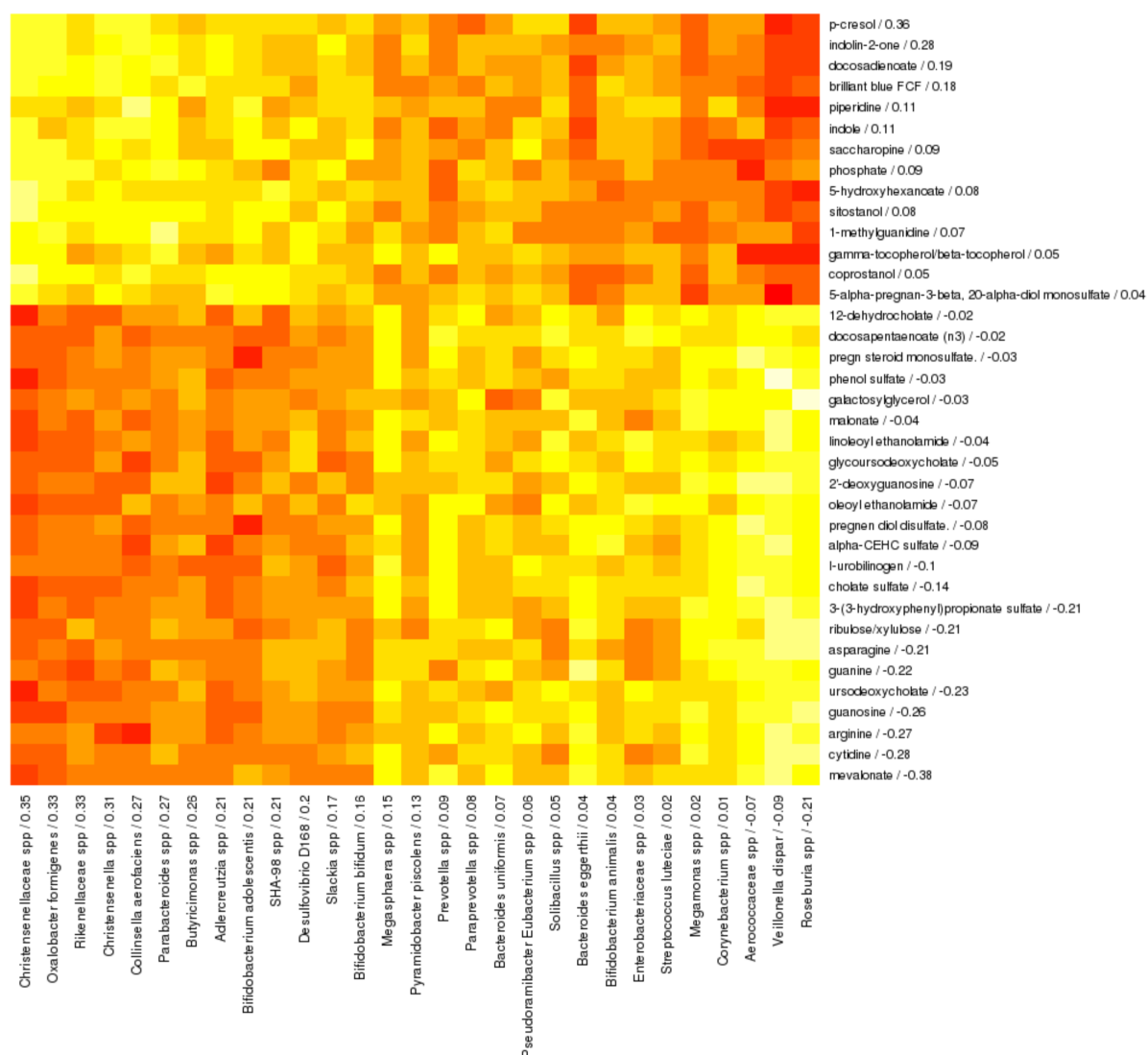


Figure 4

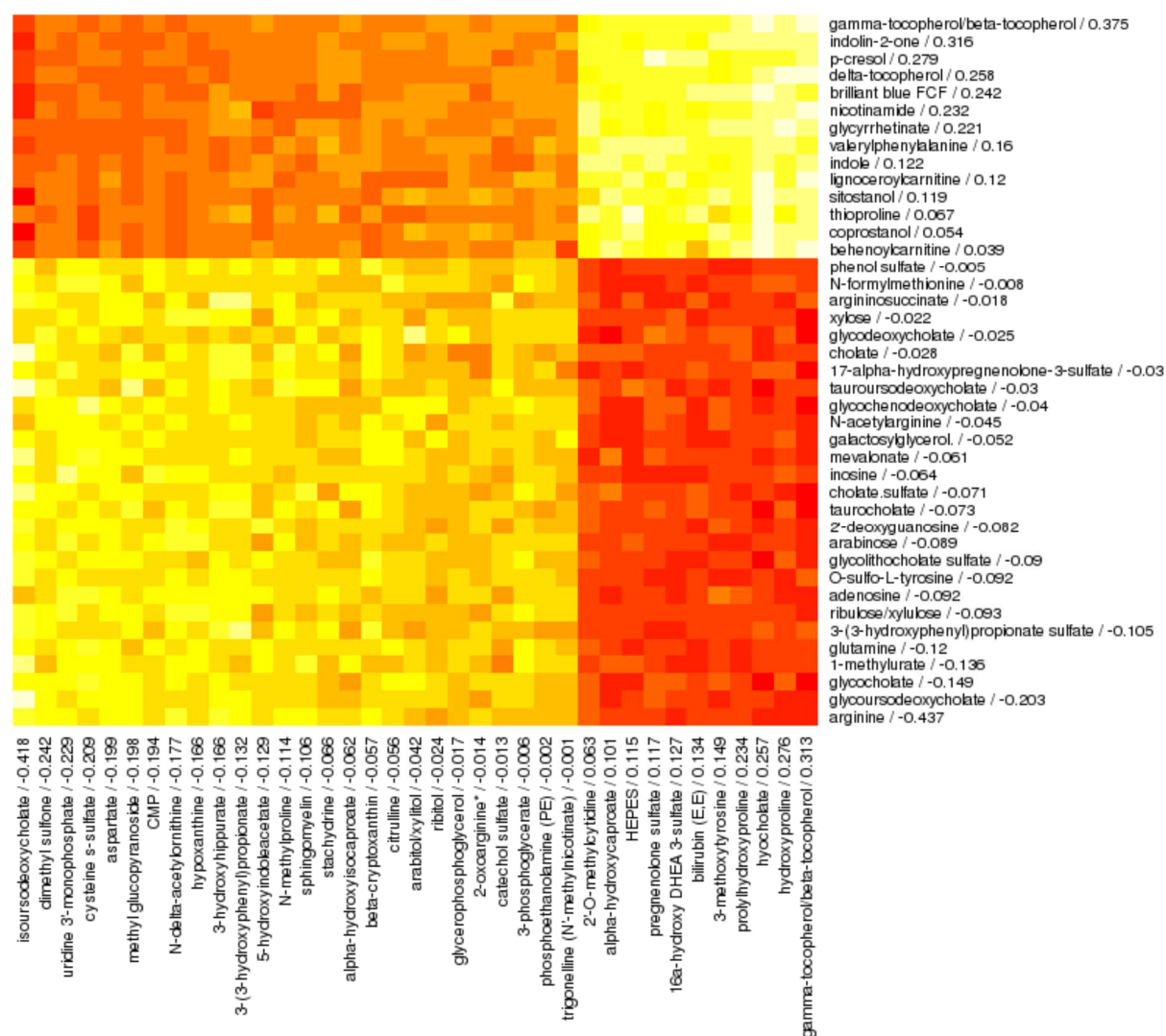


Figure 5

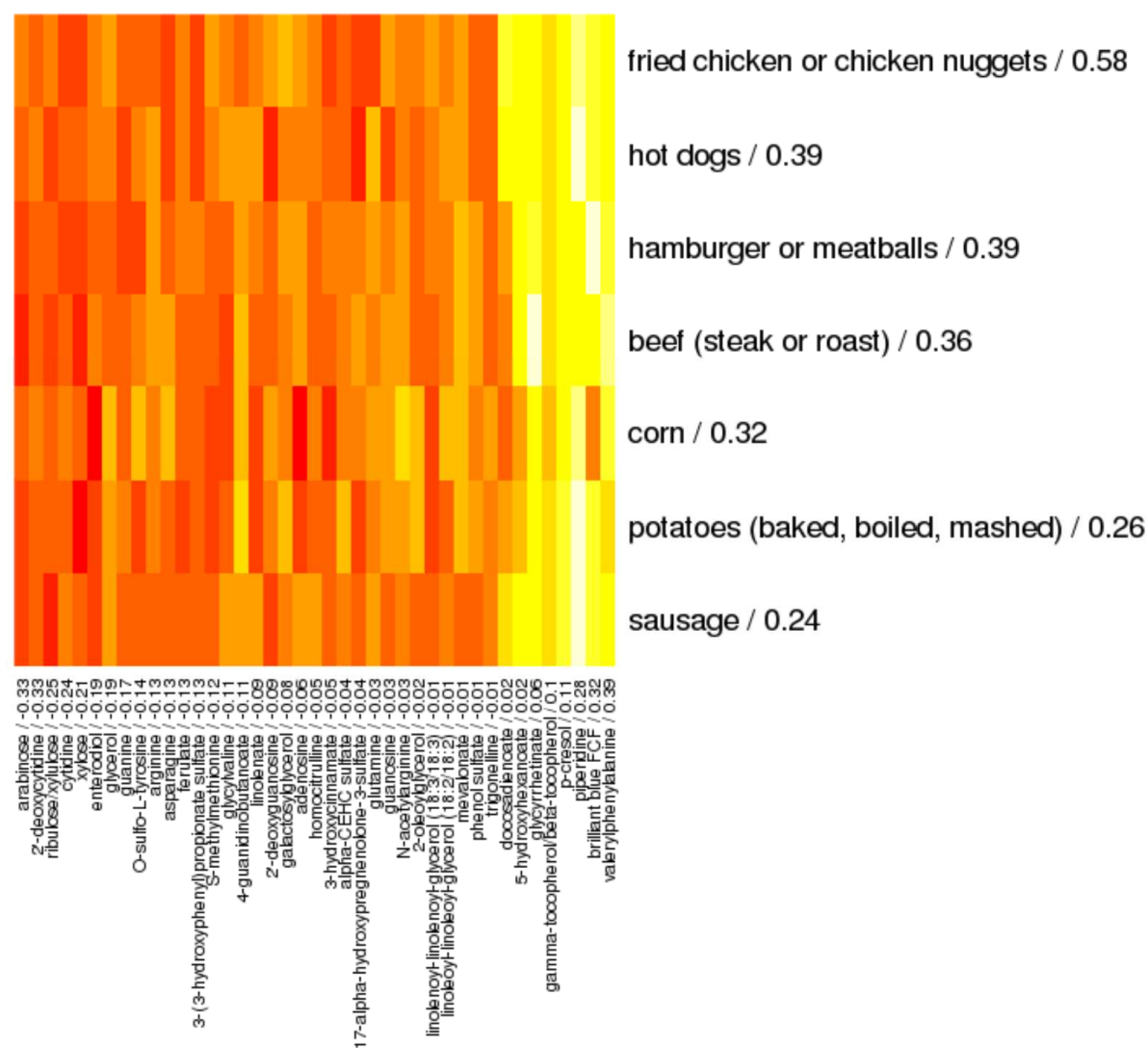


Figure 6

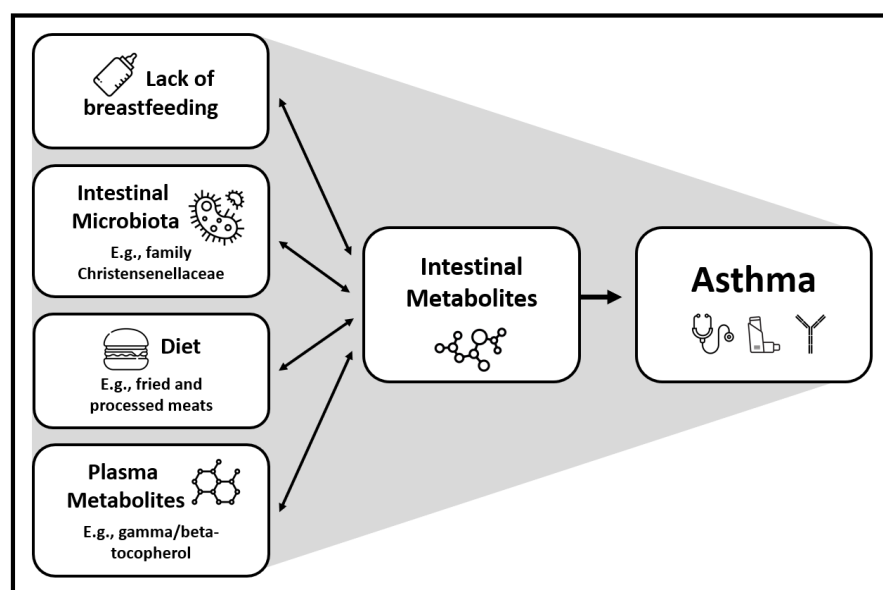


Table E1. Fecal metabolites significantly different between subjects with asthma (n=85) vs no asthma (n=276, reference group) ($p < 0.05$) at age 3 years. Metabolites are grouped by metabolic subpathway, in order of decreasing significance of the most significant metabolite in that subpathway.

Metabolite	Pathway	Subpathway	Odds Ratio (95% Confidence Interval)	p value	Including Steroid Use in Model	
					Odds Ratio (95% Confidence Interval)	p value
Docosapentaenoate (n6 DPA; 22:5n6)	Lipid	Polyunsaturated Fatty Acid	0.56 (0.40, 0.76)	0.0004	0.56 (0.43, 0.83)	0.002
Linoleate (18:2n6)	Lipid	Polyunsaturated Fatty Acid	0.64 (0.42, 0.98)	0.039	0.62 (0.39, 0.98)	0.040
Adrenate (22:4n6)	Lipid	Polyunsaturated Fatty Acid	0.72 (0.52, 1.00)	0.047	0.67 (0.47, 0.96)	0.029
Dcosapentaenoate (n3 DPA; 22:5n3)	Lipid	Polyunsaturated Fatty Acid	0.72 (0.51, 1.00)	0.050	0.63 (0.44, 0.92)	0.015
P-cresol sulfate	Amino Acid	Tyrosine Metabolism	0.56 (0.40, 0.79)	0.001	0.57 (0.39, 0.82)	0.003
O-sulfo-L-tyrosine	Amino Acid	Tyrosine Metabolism	0.67 (0.48, 0.92)	0.014	0.62 (0.43, 0.87)	0.007
Phenol sulfate	Amino Acid	Tyrosine Metabolism	0.69 (0.51, 0.93)	0.015	0.71 (0.51, 0.98)	0.043
Cis-4-hydroxycyclohexylacetic acid	Amino Acid	Tyrosine Metabolism	0.65 (0.44, 0.97)	0.033	0.60 (0.39, 0.93)	0.021
Palmitoyl-oleoyl-glycerol (16:0/18:1) [1]*	Lipid	Diacylglycerol	0.59 (0.42, 0.83)	0.002	0.67 (0.47, 0.97)	0.032
Palmitoyl-oleoyl-glycerol (16:0/18:1) [2]*	Lipid	Diacylglycerol	0.55 (0.37, 0.80)	0.002	0.61 (0.40, 0.93)	0.021
Oleoyl-oleoyl-glycerol (18:1/18:1) [1]*	Lipid	Diacylglycerol	0.64 (0.47, 0.86)	0.003	0.71 (0.51, 0.97)	0.034
Diacylglycerol (14:0/18:1, 16:0/16:1) [1]*	Lipid	Diacylglycerol	0.61 (0.43, 0.85)	0.004	0.66 (0.45, 0.94)	0.024
Oleoyl-oleoyl-glycerol (18:1/18:1) [2]*	Lipid	Diacylglycerol	0.68 (0.52, 0.88)	0.004	0.71 (0.53, 0.95)	0.019
Diacylglycerol (12:0/18:1, 14:0/16:1, 16:0/14:1) [2]*	Lipid	Diacylglycerol	0.64 (0.45, 0.88)	0.007	0.66 (0.46, 0.94)	0.025
Diacylglycerol (12:0/18:1, 14:0/16:1, 16:0/14:1) [1]*	Lipid	Diacylglycerol	0.53 (0.32, 0.82)	0.008	0.56 (0.33, 0.90)	0.025
Diacylglycerol (16:1/18:2 [2], 16:0/18:3 [1])*	Lipid	Diacylglycerol	0.71 (0.53, 0.94)	0.017	0.71 (0.52, 0.97)	0.030
Oleoyl-linoleoyl-glycerol (18:1/18:2) [1]	Lipid	Diacylglycerol	0.71 (0.52, 0.96)	0.029	0.72 (0.51, 1.00)	0.055
Linoleoyl-linolenoyl-glycerol (18:2/18:3) [2]*	Lipid	Diacylglycerol	0.74 (0.56, 0.97)	0.031	0.77 (0.57, 1.05)	0.100
17alpha-hydroxypregnenolone 3-sulfate	Lipid	Pregnenolone Steroids	0.63 (0.47, 0.85)	0.002	0.60 (0.43, 0.83)	0.002
12,13-DiHOME	Lipid	Fatty Acid, Dihydroxy	0.60 (0.41, 0.88)	0.008	0.57 (0.38, 0.86)	0.007
1-myristoylglycerol (14:0)	Lipid	Monoacylglycerol	0.56 (0.36, 0.86)	0.010	0.52 (0.31, 0.84)	0.009
Octadecenedioate (C18:1-DC)	Lipid	Fatty Acid, Dicarboxylate	0.63 (0.44, 0.90)	0.011	0.60 (0.40, 0.89)	0.011
Galactosylglycerol*	Lipid	Glycolipid Metabolism	0.65 (0.46, 0.90)	0.011	0.68 (0.47, 0.97)	0.038
2,8-quinolinediol	Xenobiotics	Food/Plant	0.70 (0.53, 0.92)	0.012	0.72 (0.53, 0.96)	0.028
Pyrraline	Xenobiotics	Food/Plant	0.64 (0.44, 0.92)	0.019	0.64 (0.43, 0.96)	0.032
Stachydrine	Xenobiotics	Food/Plant	0.70 (0.50, 0.97)	0.037	0.62 (0.43, 0.90)	0.012
3,5-dihydroxybenzoic acid	Xenobiotics	Food/Plant	0.67 (0.45, 0.98)	0.041	0.74 (0.48, 1.13)	0.170
Sitostanol	Xenobiotics	Food/Plant	1.32 (1.01, 1.76)	0.047	1.45 (1.06, 2.00)	0.022
Pyroglutamyl-leucine*	Peptide	Dipeptide	0.57 (0.36, 0.88)	0.012	0.55 (0.34, 0.88)	0.014

Linoleoyl ethanolamide	Lipid	Endocannabinoid	0.69 (0.51, 0.93)	0.014	0.67 (0.49, 0.93)	0.016
Oleoyl ethanolamide	Lipid	Endocannabinoid	0.66 (0.46, 0.94)	0.022	0.65 (0.44, 0.97)	0.035
Palmitoyl ethanolamide	Lipid	Endocannabinoid	0.67 (0.46, 0.97)	0.037	0.66 (0.43, 0.99)	0.050
Stearoyl ethanolamide	Lipid	Endocannabinoid	0.66 (0.44, 0.98)	0.042	0.64 (0.41, 0.99)	0.048
11-ketoetiocholanolone sulfate	Lipid	Androgenic Steroids	0.68 (0.50, 0.93)	0.016	0.66 (0.46, 0.93)	0.018
Xylose	Carbohydrate	Pentose Metabolism	0.66 (0.47, 0.94)	0.022	0.63 (0.42, 0.93)	0.020
Homocitrulline	Amino Acid	Urea Cycle	0.65 (0.45, 0.95)	0.026	0.61 (0.41, 0.91)	0.017
Cysteinylglycine	Amino Acid	Glutathione Metabolism	1.42 (1.04, 1.94)	0.026	1.37 (0.97, 1.93)	0.074
Ursodeoxycholate sulfate (1)	Lipid	Secondary Bile Acid	0.73 (0.54, 0.96)	0.027	0.69 (0.50, 0.94)	0.022
Cysteine s-sulfate	Amino Acid	Methionine and Cysteine Metabolism	0.58 (0.36, 0.94)	0.028	0.52 (0.30, 0.88)	0.015
Palmitoleate (16:1n7)	Lipid	Long Chain Fatty Acid	0.57 (0.34, 0.93)	0.029	0.48 (0.26, 0.83)	0.011
Lactosyl-N-behenoyl-sphingosine (d18:1/22:0)*	Lipid	Sphingolipid Metabolism	0.69 (0.48, 0.96)	0.032	0.59 (0.38, 0.86)	0.009
3-hydroxypyridine	Xenobiotics	Chemical	1.59 (1.05, 2.46)	0.033	1.76 (1.11, 2.90)	0.021
Lactosyl-N-palmitoyl-sphingosine (d18:1/16:0)	Lipid	Ceramides	0.73 (0.55, 0.97)	0.034	0.70 (0.51, 0.95)	0.025
Trigonelline (N'-methylnicotinate)	Cofactors and Vitamins	Nicotinate and Nicotinamide Metabolism	0.73 (0.54, 0.99)	0.045	0.73 (0.52, 1.02)	0.062
PAHSA (16:0/OH-18:0)	Lipid	Fatty Acid Hydroxyl Fatty Acid	0.75 (0.56, 1.00)	0.046	0.81 (0.59, 1.11)	0.180

Logistic regression analyses were adjusted for sex, race/ethnicity, study center, maternal education and gestational age.

Steroid use includes parent-reported inhaled or systemic steroid use in the three months prior to the age 3 years study visit.

*Indicates compounds with annotations that have not been officially confirmed based on a standard.

Bolded *p* values were significant after corrected for multiple testing by method of number of effective tests, as described in Methods.

Table E2. Modules of highly correlated intestinal metabolites. Ten of the 737 metabolites did not cluster into a module.*Indicates compounds with annotations that have not been officially confirmed based on a standard.

Module 1	Module 5	Module 8
1,2-dilinolenoyl-galactosylglycerol (18:3/18:3)*	tauroursodeoxycholate	guanosine
1-palmitoyl-2-linoleoyl-digalactosylglycerol (16:0/18:2)*	taurocholate	adenosine
1,2-dilinoleoyl-galactosylglycerol (18:2/18:2)*	taurine	2'-deoxyguanosine
1,2-dilinoleoyl-digalactosylglycerol (18:2/18:2)*	N-acetyltaurine	inosine
1-linoleoyl-2-linolenoyl-galactosylglycerol (18:2/18:3)*	taurochenodeoxycholate	arginine
1-palmitoyl-2-linoleoyl-galactosylglycerol (16:0/18:2)*	Module 6	3-hydroxybenzoate
pheophorbide A	vanillate	asparagine
Module 2	S-methylmethionine	argininosuccinate
palmitoyl-oleoyl-glycerol (16:0/18:1) [2]*	glycoursodeoxycholate	mevalonate
oleoyl-oleoyl-glycerol (18:1/18:1) [2]*	7-ketodeoxycholate	cytidine
diacylglycerol (14:0/18:1, 16:0/16:1) [1]*	glycolithocholate sulfate*	2'-deoxyadenosine
oleoyl-oleoyl-glycerol (18:1/18:1) [1]*	biliverdin	2'-deoxycytidine
palmitoyl-oleoyl-glycerol (16:0/18:1) [1]*	glycochenodeoxycholate	guanine
diacylglycerol (16:1/18:2 [2], 16:0/18:3 [1])*	3b-hydroxy-5-cholenoic acid	adenine
diacylglycerol (12:0/18:1, 14:0/16:1, 16:0/14:1) [2]*	glycocholate	4-guanidinobutanoate
diacylglycerol (12:0/18:1, 14:0/16:1, 16:0/14:1) [1]*	7-ketolithocholate	7-methylguanine
oleoyl-linoleoyl-glycerol (18:1/18:2) [1]	3-hydroxyhexanoate	agmatine
linoleoyl-linolenoyl-glycerol (18:2/18:3) [2]*	cholate	xanthosine
linoleoyl-linoleoyl-glycerol (18:2/18:2) [1]*	2-hydroxyadipate	homoarginine
linoleoyl-linoleoyl-glycerol (18:2/18:2) [2]*	sucrose	2-isopropylmalate
palmitoyl-linolenoyl-glycerol (16:0/18:3) [2]*	N-glycolylneuramate	uridine
palmitoyl-linoleoyl-glycerol (16:0/18:2) [1]*	N-acetylproline	8-hydroxyguanine
palmitoleoyl-linoleoyl-glycerol (16:1/18:2) [1]*	syringic acid	2,3-dihydroxyisovalerate
linolenoyl-linolenoyl-glycerol (18:3/18:3) [2]*	chenodeoxycholate	2-oxoarginine*
oleoyl-linoleoyl-glycerol (18:1/18:2) [2]	genistein	4-hydroxybenzoate
diacylglycerol (14:0/18:1, 16:0/16:1) [2]*	12-dehydrocholate	nicotianamine
oleoyl-linolenoyl-glycerol (18:1/18:3) [2]*	tetrahydrocortisol	malonate
linolenoyl-linolenoyl-glycerol (18:3/18:3) [1]*	cortolone	2'-deoxyinosine
13-HODE + 9-HODE	bilirubin (Z,Z)	N6-carbamoylthreonyladenosine
tartarate	vanillactate	thymidine
palmitoyl-linoleoyl-glycerol (16:0/18:2) [2]*	digalacturonic acid	5-methyluridine (ribothymidine)
13-HOTrE	4-hydroxyphenylacetate	2'-deoxyuridine
Module 3	quinate	Module 9
21-hydroxypregnenolone monosulfate (1)	taurocholenate sulfate	5-(2-Hydroxyethyl)-4-methylthiazole
pregn steroid monosulfate*	glycerophosphoserine*	xanthine
pregnen-diol disulfate*	3-dehydrocholate	glutamate, gamma-methyl ester
dehydroisoandrosterone sulfate (DHEA-S)	glycocholenate sulfate*	diaminopimelate
pregnenolone sulfate	saccharin	thymine
androstenediol (3beta,17beta) monosulfate (1)	7,12-diketolithocholate	nicotinate ribonucleoside
5alpha-pregnan-3beta,20alpha-diol disulfate	3-hydroxyisobutyrate	fumarate
androstenediol (3beta,17beta) disulfate (2)	maltotetraose	gamma-glutamyl-2-aminobutyrate
androstenediol (3beta,17beta) disulfate (1)	maltitol/lactitol/cellobiotol/palatinol	glutamyl-meso-diaminopimelate
Module 4	Module 7	uracil
lactosyl-N-behenoyl-sphingosine (d18:1/22:0)*	carotene diol (2)	1-methyladenine
lactosyl-N-palmitoyl-sphingosine (d18:1/16:0)	carotene diol (3)	hydroxymethylpyrimidine
lactosyl-N-nervonoyl-sphingosine (d18:1/24:1)*	beta-cryptoxanthin	flavin adenine dinucleotide (FAD)
palmitoyl dihydrosphingomyelin (d18:0/16:0)*	carotene diol (1)	N-acetylmuramate
palmitoyl sphingomyelin (d18:1/16:0)	2-oxindole-3-acetate	pseudouridine
glycosyl-N-palmitoyl-sphingosine (d18:1/16:0)		sedoheptulose
glycosyl-N-stearoyl-sphingosine (d18:1/18:0)		malate
7-hydroxycholesterol (alpha or beta)		

Table E2 (Continued)

Module 10	Module 15	Module 19
xylose	p-cresol sulfate	stachydrine
galactosylglycerol*	O-sulfo-L-tyrosine	2,4-dihydroxyhydrocinnamate
ferulate	phenol sulfate	epicatechin
ribulose/xylulose	2,8-quinolinediol	apigenin
arabinose	alpha-CEHC sulfate	ponciretin
maltose	4-methylcatechol sulfate	salicylate
glucose	furaneol sulfate	3-hydroxy-3-methylglutarate
glucuronate	catechol sulfate	hesperetin
fructose	pyridoxine (Vitamin B6)	naringenin
galacturonate	2-aminophenol sulfate	chiro-inositol
mannose	homovanillate sulfate	2,4,6-trihydroxybenzoate
dihydroferulic acid	Module 16	2-aminophenol
maltotriose	eicosanoylsphingosine (d20:1)*	Module 20
Module 11	N-acetylsphingosine	p-cresol
oleoyl ethanolamide	N-2-hydroxypalmitoyl-sphingosine (d18:1/16:0(2OH))	N-acetylalanine
12,13-DiHOME	hexadecaspheganine (d16:0)*	valerylphenylalanine
linoleoyl ethanolamide	ceramide (d18:1/14:0, d16:1/16:0)*	N-acetylphenylalanine
stearoyl ethanolamide	heptadecaspheganine (d17:1)	3-phenylpropionate (hydrocinnamate)
palmitoyl ethanolamide	N-palmitoyl-sphingosine (d18:1/16:0)	indole
arachidoyl ethanolamide (20:0)*	N-butyryl-sphingosine (d18:1/4:0)	N-acetylvaline
5-methylthioadenosine (MTA)	N-palmitoyl-sphinganine (d18:0/16:0)	phenylpropionylglycine
erucoyl ethanolamide (22:1)*	sphingosine	valerylglycine
margaroyl ethanolamide*	N-stearoyl-sphingosine (d18:1/18:0)*	2-hydroxyphenylacetate
9,10-DiHOME	hexadecaspheganine (d16:1)*	N-acetylleucine
daidzein	sphinganine	diethanolamine
Module 12	Module 17	phenylacetate
caprate (10:0)	pterin	2-acetamidobutanoate
myristoleate (14:1n5)	pimelate (heptanedioate)	phenylacetylphenylalanine
5-dodecenoate (12:1n7)	undecanedioate	N-acetyltyrosine
laurate (12:0)	suberate (octanedioate)	2-methylbutyrylphenylalanine
myristate (14:0)	azelate (nonanedioate)	N-propionylalanine
phytanate	adipate	N-acetylisoleucine
undecanoate (11:0)	diglycerol	N-methylalanine
pristanate	pyridoxal	tyrosol
Module 13	sebacate (decanedioate)	butyrylglycine
behenoylcarnitine (C22)*	O-acetylhomoserine	3-(2-hydroxyphenyl)propionate
piperine	3-methylglutaconate	N-formylanthranilic acid
arachidoylcarnitine (C20)*	6-oxopiperidine-2-carboxylate	propionylglycine
nervonoylcarnitine (C24:1)*	3-hydroxysuberate	norvaline
lignoceroylcarnitine (C24)*	N-methylpipercolate	Module 21
behenoyl ethanolamide (22:0)*	N-trimethyl 5-aminovalerate	1-stearoyl-GPE (18:0)
cerotoylcarnitine (C26)*	glutarate (pentanedioate)	1-stearoyl-GPG (18:0)
2-methylmalonylcarnitine (C4-DC)	3-hydroxybutyrylcarnitine (2)	1-palmitoyl-GPG (16:0)*
Module 14	methylsuccinate	1-stearoyl-GPS (18:0)*
2-piperidinone	dodecanedioate	1-palmitoyl-GPE (16:0)
trimethylamine N-oxide	Module 18	glycerophosphoethanolamine
2-aminoadipate	3-(3-hydroxyphenyl)propionate sulfate	1-palmitoyl-GPI (16:0)
5-aminovalerate	3-(3-hydroxyphenyl)propionate	1-(1-enyl-palmitoyl)-GPE (P-16:0)*
phenethylamine	3-hydroxycinnamate	1-stearoyl-GPI (18:0)
N-acetyltryptophan	4-hydroxycinnamate	1-(1-enyl-stearoyl)-GPE (P-18:0)*
	3-(4-hydroxyphenyl)propionate	1-oleoyl-GPE (18:1)

Table E2 (Continued)

Module 22	Module 26	Module 26 (Continued)
3-hydroxypyridine	cysteine s-sulfate	proline
3,5-dihydroxybenzoic acid	homocitrulline	1-methyl-beta-carboline-3-carboxylic acid
3-methyladipate	cis-4-hydroxycyclohexylacetic acid	aspartate
N-acetylcysteine	1-methylurate	picolinate
quinolinate	N-acetylglucosaminylasparagine	cis-urocanate
dimethylmalonic acid	pyr-leu*	N-acetylserine
3-methylglutarate/2-methylglutarate	N-acetylarginine	4-acetamidobenzoate
biotin	levulinate (4-oxovalerate)	cysteine sulfinic acid
N-methylproline	glutamine	riboflavin (Vitamin B2)
enterolactone	glycylvaline	gamma-glutamyl-epsilon-lysine
carboxyethyl-GABA	N-acetylasparagine	mevalonolactone
hydantoin-5-propionic acid	serine	N-acetylmethionine sulfoxide
pipecolate	cystine	glutamate
serotonin	tryptophan	N-acetylhistidine
ethylmalonate	N-acetylglutamine	10-undecenoate (11:1n1)
2,3-dimethylsuccinate	3-hydroxy-2-ethylpropionate	ribose
Module 23	o-Tyrosine	pyridoxamine
1-methylguanidine	biocytin	N2-acetyllysine
ectoine	glycylisoleucine	beta-hydroxyisovalerate
1-methylimidazoleacetate	3-sulfo-L-alanine	N-delta-acetylorithine
3-methylhistidine	N-formylmethionine	2-methylserine
N-acetyl-3-methylhistidine*	N-alpha-acetylorithine	alanine
anserine	histidine	Module 27
Module 24	cysteine	glycodeoxycholate
gluconate	indoleacetate	linolenoylcarnitine (C18:3)*
tricarballoylate	threonine	deoxycholate
gulonate*	N6-formyllysine	dehydrolithocholate
N-acetylaspartate (NAA)	pyridoxate	6-oxolithocholate
protoporphyrin IX	kynurenine	lithocholate
phosphate	pyroglutamine*	Module 28
3-carboxyadipate	phenylalanine	kynurenate
gamma-glutamylglutamate	ornithine	N-acetylglutamate
glycerol 3-phosphate	N6-acetyllysine	oleanolate
galactonate	lysine	gamma-glutamyl-alpha-lysine
methylphosphate	acetylcarnitine (C2)	gamma-glutamyltyrosine
aconitate [cis or trans]	methionine sulfoxide	gamma-glutamylmethionine
2-methylcitrate/homocitrate	citrulline	4-hydroxycyclohexylcarboxylic acid
oxalate (ethanedioate)	N-propionylmethionine	gamma-glutamylglycine
flavin mononucleotide (FMN)	N-acetylmethionine	5-oxoproline
citrate	tyrosine	gamma-glutamylleucine
Module 25	methionine sulfone	gamma-glutamylglutamine
enterodiol	leucine	gamma-glutamylalanine
linoleate (18:2n6)	isoleucine	gamma-glutamylthreonine
2-oleoylglycerol (18:1)	beta-guanidinopropanoate	gamma-glutamylphenylalanine
glycerol	spermidine	gamma-glutamylisoleucine*
linolenate [alpha or gamma; (18:3n3 or 6)]	N-acetylglycine	Module 29
sinapate	glycine	piperidine
1-oleoylglycerol (18:1)	valine	cadaverine
1-linoleoylglycerol (18:2)	methionine	histamine
2-linoleoylglycerol (18:2)	N-methyl-GABA	N-acetyl-cadaverine
caffeate	allo-threonine	1-methylhistamine
1-linolenoylglycerol (18:3)	2'-O-methyluridine	Tyramine
1-palmitoleoylglycerol (16:1)*	succinimide	
choline phosphate	trans-urocanate	

Table E2 (Continued)

Module 30	Module 33	Module 39
N1-Methyl-2-pyridone-5-carboxamide	glycyrhettinate	3,7-dimethylurate
alpha-CEHC	LAHSA (18:2/OH-18:0)*	7-methylurate
phytosphingosine	OAHA (18:1/OH-18:0)*	(R)-salsolinol
N6-carboxyethyllysine	PAHSA (16:0/OH-18:0)	3-methylurate*
tryptamine	alpha-tocotrienol	3-methylxanthine
ursodeoxycholate	gamma-tocotrienol	theobromine
ursocholate	Module 34	m-tyramine
deoxycarnitine	trigonelline (N'-methylnicotinate)	Module 40
hyocholate	4-acetamidobutanoate	palmitoleate (16:1n7)
4-hydroxyhippurate	N2,N5-diacetylornithine	oleate/vaccenate (18:1)
androsterone sulfate	2-pyrrolidinone	docosadienoate (22:2n6)
gamma-CEHC	Module 35	N-palmitoylserine
isoursodeoxycholate	17alpha-hydroxypregnenolone 3-sulfate	hexadecadienoate (16:2n6)
N6-carboxymethyllysine	ursodeoxycholate sulfate (1)	palmitate (16:0)
2-hydroxydecanoate	5alpha-pregnan-3beta,20beta-diol monosulfate (1)	eicosenoate (20:1)
Module 31	11-ketoetiocolanolone sulfate	trans-nonadecenoate (tr 19:1)*
3-hydroxyphenylacetate	cholate sulfate	10-hydroxystearate
lactobionate	pregnanolone/allopregnanolone sulfate	15-methylpalmitate
2-(4-hydroxyphenyl)propionate	l-urobilinogen	N-palmitoylglycine
argininate*	bilirubin (E,Z or Z,E)*	dihomo-linoleate (20:2n6)
myo-inositol	D-urobilin	13-methylmyristate
cytosine	Module 36	oleoyl-arachidonoyl-glycerol (18:1/20:4) [1]*
tryptophan betaine	isovalerylglutamate	nonadecanoate (19:0)
nicotinate	2-methylbutyrylglutamate	margarate (17:0)
3-deoxyoctulosonate	isobutyrylglutamate	17-methylstearate
gamma-glutamylhistidine	N-formylphenylalanine	arachidate (20:0)
indolepropionate	Module 37	erucate (22:1n9)
xanthurenate	N-acetyl-beta-glucosaminylamine	stearate (18:0)
1-methylxanthine	tryptophylglutamate	Module 41
pantothenate	lysylleucine	urate
arabitol/xylitol	uridine-2',3'-cyclic monophosphate	S-1-pyrroline-5-carboxylate
pyruvate	1-methylnicotinamide	guanidininosuccinate
mannitol/sorbitol	leucylglutamate	formiminoglutamate
hypoxanthine	alanylleucine	N-methylleucine
glycerophosphoglycerol	leucylglutamine*	p-hydroxybenzaldehyde
gentisate	threonylphenylalanine	4-ureidobutyrate
lactate	glycylleucine	N6-methyladenosine
Module 32	histidylalanine	allantoin
2,3-dihydroxy-2-methylbutyrate	R-mevalonate 5-diphosphate	phenylpyruvate
cysteinylliglutamate	valylleucine	3-ketosphinganine
3-hydroxybutyrate (BHBA)	isoleucylglutamate	4-imidazoleacetate
N6,N6,N6-trimethyllysine	valylglutamine	alpha-ketobutyrate
6-hydroxynicotinate	phenylalanylalanine	N-methyltryptophan
3,4-dihydroxyphenylacetate	phenylalanylliglutamate	3-methyl-2-oxobutyrate
N-acetylhistamine	leucylalanine	4-hydroxyphenylpyruvate
3-hydroxyvalerate	valylglutamate	4-methylthio-2-oxobutanoate
N1,N12-diacetylspermine	glutaminylleucine	N-acetylserotonin
N(1)-acetylspermine	nicotinamide riboside	retinol (Vitamin A)
S-carboxymethyl-L-cysteine	Module 38	3-methyl-2-oxovalerate
N(1)-acetylspermidine	3-ureidopropionate	4-methyl-2-oxopentanoate
N-acetylputrescine	3-aminoisobutyrate	indole-3-carboxylic acid
N-monomethylarginine	5,6-dihydrothymine	alpha-ketoglutarate
putrescine	beta-alanine	methyl indole-3-acetate
dimethylarginine (SDMA + ADMA)	N-acetyl-beta-alanine	

Table E2 (Continued)

Module 42	Module 47	Module 52
palmitoylcarnitine (C16)	pyrraline	indolin-2-one
stearoylcarnitine (C18)	phenylacetylglutamine	nicotinamide
oleoylcarnitine (C18:1)	guanidinoacetate	sitostanol
myristoylcarnitine (C14)	5-hydroxylysine	brilliant blue FCF (blue 1)
palmitoleoylcarnitine (C16:1)*	N-acetylpyrraline	coprostanol
3-methylglutaryl carnitine (2)	N-(2-furoyl)glycine	sucralose
succinyl carnitine (C4-DC)	creatine	gamma-tocopherol/beta-tocopherol
linoleoylcarnitine (C18:2)*	hippurate	saccharopine
carnitine	dimethylglycine	harmaline
Module 43	N-methylhydantoin	5-hydroxyhexanoate
thioproline	trans-4-hydroxyproline	delta-tocopherol
imidazole propionate	N-acetyl-1-methylhistidine*	2-hydroxymyristate
2-keto-3-deoxy-gluconate	sarcosine	5alpha-pregnan-3beta,20alpha-diol monosulfate (1)
ribitol	acisoga	tyramine O-sulfate
2-hydroxy-3-methylvalerate	betaine	orotate
succinylglutamine	carosine	2-hydroxypalmitate
succinate	creatinine	dihydroorotate
alpha-hydroxyisocaproate	benzoate	pentadecanoate (15:0)
3-(4-hydroxyphenyl)lactate	1-methylhistidine	4-cholesten-3-one
phenylacetyl glycine	Module 48	2-hydroxybehenate
alpha-hydroxyisovalerate	N-acetylneuraminate	2-hydroxystearate
2-hydroxyglutarate	N-acetylglucosamine/N-acetylgalactosamine	L-urobilin
phenyllactate (PLA)	erythrose	N-carbamoylaspartate
2-hydroxybutyrate/2-hydroxyisobutyrate	fucose	N-methylphenylalanine
N-acetylcitrulline	Module 49	thiamin (Vitamin B1)
N-acetylthreonine	1-myristoylglycerol (14:0)	triethanolamine
indolelactate	2-palmitoylglycerol (16:0)	alpha-tocopherol
phenylacetylglutamate	1-palmitoylglycerol (16:0)	Module 53
Module 44	1-pentadecanoylglycerol (15:0)	cholesterol
3-hydroxylaurate	1-margaroylglycerol (17:0)	hexadecanedioate
3-hydroxydecanoate	Module 50	campesterol
3-hydroxymyristate	1-palmitoyl-GPC (16:0)	stigmasterol
3-hydroxyoctanoate	1-palmitoyl-2-oleoyl-GPC (16:0/18:1)	desmosterol
Module 45	1-stearoyl-GPC (18:0)	beta-sitosterol
theophylline	choline	3-hydroxysebacate
paraxanthine	1-oleoyl-GPC (18:1)	Module 54
1,3-dimethylurate	1-palmitoyl-2-linoleoyl-GPC (16:0/18:2)	heptanoate (7:0)
5-acetylamino-6-formylamino-3-methyluracil	glycerophosphoinositol*	caproate (6:0)
1,7-dimethylurate	1-linoleoyl-GPC (18:2)	valerate
1,3,7-trimethylurate	1-linoleoyl-GPE (18:2)*	isovalerate
5-acetylamino-6-amino-3-methyluracil	Module 51	caprylate (8:0)
Module 46	ribonate	isocaproate
docosapentaenoate (n6 DPA; 22:5n6)	N-acetyl-aspartyl-glutamate (NAAG)	
octadecenedioate (C18:1-DC)	glycerate	
docosapentaenoate (n3 DPA; 22:5n3)	4-hydroxyglutamate	
docosahexaenoate (DHA; 22:6n3)	imidazole lactate	
dihomo-linolenate (20:3n3 or n6)	sulfate*	
eicosapentaenoate (EPA; 20:5n3)	arabonate/xylonate	
adrenate (22:4n6)	maleate	
arachidonate (20:4n6)	fucitol	
mead acid (20:3n9)	erythronate*	
	threonate	

Table E3. Members of modules of highly-correlated metabolites that were inversely associated with asthma at age 3 years.

Diacylglycerol Module	Endocannabinoid Module	Polyunsaturated Fatty Acids Module
palmitoyl-oleoyl-glycerol (16:0/18:1) [2]*	5-methylthioadenosine (MTA)	Docosapentaenoate (n6 DPA)
oleoyl-oleoyl-glycerol (18:1/18:1) [2]*	oleoyl ethanolamide	Octadecenedioate
diacylglycerol (14:0/18:1, 16:0/16:1) [1]*	linoleoyl ethanolamide	Docosapentaenoate (n3 DPA)
oleoyl-oleoyl-glycerol (18:1/18:1) [1]*	stearoyl ethanolamide	Docosahexaenoate (DHA)
palmitoyl-oleoyl-glycerol (16:0/18:1) [1]*	palmitoyl ethanolamide	Dihomo-linolenate (n3 or n6)
diacylglycerol (16:1/18:2 [2], 16:0/18:3 [1])*	arachidoyl ethanolamide (20:0)*	Eicosapentaenoate (EPA)
diacylglycerol (12:0/18:1, 14:0/16:1, 16:0/14:1) [2]*	erucoyl ethanolamide (22:1)*	Adrenate
diacylglycerol (12:0/18:1, 14:0/16:1, 16:0/14:1) [1]*	margaroyl ethanolamide*	Arachidonate
oleoyl-linoleoyl-glycerol (18:1/18:2) [1]	9,10-DiHOME	Mead acid
linoleoyl-linolenoyl-glycerol (18:2/18:3) [2]*	12,13-DiHOME	
linoleoyl-linoleoyl-glycerol (18:2/18:2) [1]*	daidzein	
linoleoyl-linoleoyl-glycerol (18:2/18:2) [2]*		
palmitoyl-linolenoyl-glycerol (16:0/18:3) [2]*		
palmitoyl-linoleoyl-glycerol (16:0/18:2) [1]*		
palmitoleoyl-linoleoyl-glycerol (16:1/18:2) [1]*		
linolenoyl-linolenoyl-glycerol (18:3/18:3) [2]*		
oleoyl-linoleoyl-glycerol (18:1/18:2) [2]		
diacylglycerol (14:0/18:1, 16:0/16:1) [2]*		
oleoyl-linolenoyl-glycerol (18:1/18:3) [2]*		
linolenoyl-linolenoyl-glycerol (18:3/18:3) [1]*		
13-HODE + 9-HODE		
Tartarate		
palmitoyl-linoleoyl-glycerol (16:0/18:2) [2]*		
13-HOTrE		

*Indicates compounds with annotations that have not been officially confirmed based on a standard.

Table E4. Asthma-associated fecal metabolites associated with exclusive breastfeeding for the first 4 months of life.

Metabolite	Association with Breastfeeding		Association with Asthma		Mediation Analysis Results*	
	Beta (95% CI)	p value	OR (95% CI)	p value	P value for indirect effect	Estimated proportion mediated
Xylose	0.23 (0.04, 0.42)	0.02	0.66 (0.47, 0.94)	0.02	0.04	11.1%
P-cresol sulfate	0.25 (0.05, 0.45)	0.01	0.56 (0.40, 0.79)	0.001	0.02	17.3%
O-sulfo-L-tyrosine	0.24 (0.03, 0.45)	0.02	0.67 (0.48, 0.92)	0.01	0.06	10.9%
17 α -hydroxypregnenolone 3-sulfate	0.27 (0.04, 0.49)	0.02	0.63 (0.47, 0.85)	0.002	0.03	15.5%
Galactosylglycerol	0.27 (0.06, 0.48)	0.01	0.65 (0.46, 0.90)	0.01	0.04	12.6%
linoleoyl-linolenoyl-glycerol (18:2/18:3) [2]	0.32 (0.08, 0.56)	0.01	0.74 (0.55, 0.97)	0.03	0.13	9.0%
diacylglycerol (14:0/18:1, 16:0/16:1 [1])	0.32 (0.13, 0.51)	0.001	0.61 (0.43, 0.85)	0.004	0.01	20.6%
diacylglycerol (16:1/18:2 [2], 16:0/18:3 [1])	0.30 (0.08, 0.53)	0.01	0.71 (0.53, 0.94)	0.02	0.08	11.0%
diacylglycerol (12:0/18:1, 14:0/16:1, 16:0/14:1) [1]	0.31 (0.12, 0.50)	0.001	0.53 (0.32, 0.82)	0.01	0.01	22.3%
diacylglycerol (12:0/18:1, 14:0/16:1, 16:0/14:1) [2]	0.44 (0.12, 0.50)	0.0004	0.61 (0.43, 0.85)	0.004	0.03	23.2%
oleoyl-oleoyl-glycerol (18:1/18:1) [2]	0.27 (0.02, 0.52)	0.03	0.68 (0.52, 0.88)	0.004	0.05	13.1%
palmitoyl-oleoyl-glycerol (16:0/18:1) [1]	0.21 (0.01, 0.42)	0.04	0.59 (0.42, 0.83)	0.002	0.05	12.8%

Beta effect estimates are from linear regression analyses of the association of breastfeeding with asthma-associated metabolites (n = 45 metabolites), adjusted for sex, race/ethnicity, study center, maternal education, gestational age, perinatal antibiotics, mode of delivery, and VDAART trial treatment assignment. Odds ratios for associations with asthma are from logistic regression analyses adjusted for sex, race/ethnicity, study center, maternal education and gestational age.

*Results of mediation analyses to evaluate statistical significance of the indirect effect of breastfeeding on asthma that is mediated by each metabolite; and to estimate the proportion of the association of breastfeeding with asthma that is mediated by each metabolite.

Table E5. Canonical loadings for bacterial taxa with non-zero loadings in sparse canonical correlation analysis of intestinal metabolites with intestinal bacterial taxa. Loadings for the analysis including individual metabolites are extremely similar to those for the analysis including groups of highly correlated metabolites (Pearson rho of canonical loadings = 0.99, $p < 0.0001$, $n = 273$ subjects).

Bacterial Taxa	Canonical Loading	
	Individual Metabolites	Modules of Correlated Metabolites
<i>Roseburia</i> spp	-0.21	-0.22
<i>Veillonella</i> dispar	-0.09	-0.11
Aerococcaceae spp	-0.07	-0.09
<i>Corynebacterium</i> spp	0.01	0.01
<i>Streptococcus</i> luteciae	0.02	0.01
<i>Megamonas</i> spp	0.02	-0.002
Enterobacteriaceae spp	0.03	0.02
<i>Bacteroides</i> eggerthii	0.04	0.03
<i>Bifidobacterium</i> animalis	0.04	0.03
<i>Solibacillus</i> spp	0.05	0.04
<i>Pseudoramibacter</i> <i>Eubacterium</i> spp	0.06	0.06
<i>Bacteroides</i> uniformis	0.07	0.06
<i>Paraprevotella</i> spp	0.08	0.07
<i>Prevotella</i> spp	0.09	0.08
<i>Pyramidobacter</i> piscolens	0.13	0.14
<i>Megasphaera</i> spp	0.15	0.12
<i>Bifidobacterium</i> bifidum	0.16	0.14
<i>Slackia</i> spp	0.17	0.17
<i>Desulfovibrio</i> D168	0.2	0.18
<i>Adlercreutzia</i> spp	0.21	0.22
<i>Bifidobacterium</i> adolescentis	0.21	0.2
SHA-98 spp	0.21	0.24
<i>Butyricimonas</i> spp	0.26	0.24
<i>Collinsella</i> aerofaciens	0.27	0.24
<i>Parabacteroides</i> spp	0.27	0.24
<i>Christensenella</i> spp	0.31	0.31
<i>Oxalobacter</i> formigenes	0.33	0.32
Rikenellaceae spp	0.33	0.35
Christensenellaceae spp	0.35	0.40

Table E6. Pearson rhos for correlations of plasma lipid summary measures with intestinal lipid metabolite module eigenvalues. Intestinal lipid module eigenvalues were generated by network analysis to identify modules of highly-correlated intestinal metabolites. Plasma lipid summary measures were generated by summing relative abundances of total PUFA (13 metabolites), omega-3 PUFA (5 metabolites), omega-6 PUFA (6 metabolites), endocannabinoid (6 metabolites) and diacylglycerol (19 metabolites) metabolites.

Intestinal Metabolite Module	Plasma Metabolite Summary Measure	Number of Subjects	Pearson Rho	<i>p</i> value
PUFA	Total PUFA	156	-0.02	0.77
PUFA	Omega-3 PUFA	156	0.01	0.86
PUFA	Omega-6 PUFA	156	-0.06	0.48
Diacylglycerol	Diacylglycerol	156	0.04	0.61
Endocannabinoid	Endocannabinoid	233	-0.06	0.40

Abbreviations: PUFA – polyunsaturated fatty acid

Table E7. Canonical loadings for foods with non-zero loadings in sparse canonical correlation analysis of intestinal metabolites with diet at age 3 years. Loadings for the analysis including individual metabolites are similar to those for the analysis including groups of highly correlated metabolites (Pearson rho of canonical loadings = 0.86, $p = 0.01$, $n = 338$ subjects).

Food	Canonical Loading	
	Individual Metabolites	Modules of Correlated Metabolites
Fried chicken or chicken nuggets	0.58	0.54
Hamburger or meatballs	0.39	0.43
Hot dogs	0.39	0.48
Beef (steak or roast)	0.36	0.45
Corn	0.32	0.22
Potatoes (baked, boiled, mashed)	0.26	0.14
Sausage	0.24	0.15

SUPPLEMENTAL METHODS

Study Design

Stool samples were analyzed from offspring of participants in VDAART, a randomized, double-blind, placebo-controlled trial of Vitamin D supplementation during pregnancy for prevention of asthma and other allergic disease in offspring (NCT00920621)(S1, S2). VDAART participants were recruited during the 10th to 18th weeks of pregnancy from study centers in Boston, St. Louis and San Diego, United States. Participants were randomized to daily 4,400 IU (treatment arm) or 400 IU (usual care) of vitamin D during pregnancy. VDAART offspring were at elevated risk for developing asthma and other allergic diseases, as either their biologic mother or father had a history of asthma, eczema or allergic rhinitis by study design. All participants were non-smokers. The study protocol was approved by the institutional review boards at each participating institution and at Brigham and Women's Hospital. All participants provided written informed consent. A flow diagram is displayed in **Figure 1**.

Outcome Ascertainment and Other Variables

The asthma outcome was based on parental report of physician diagnosis of asthma in the child's first 3 years of life. Parents were queried about asthma diagnosis regularly for the first 3 years of life by telephone every 3 months and at annual in-person visits. Other characteristics analyzed included child sex, race/ethnicity, study center, maternal education, and household income. Data collected at birth included maternal antibiotics during delivery or child antibiotics after delivery, mode of delivery and gestational age. Parental questionnaire responses were used to determine whether the child was exclusively breastfed for the first 4 months of life,

presence of a pet dog in the home during infancy, day care attendance, diagnosis of eczema by a health care provider, and whether inhaled (by inhaler or nebulizer) or systemic (oral or injected) steroids had been used in the 3 months prior to turning 3 years old. Blood samples collected at age 3 years were used to measure total IgE and serum specific IgE to common food and environmental allergens: egg white, milk, peanut, soybean, wheat, walnut, *Alternaria alternata*, *Dermatophagoides farinae*, *Dermatophagoides pteronyssinus*, German cockroach, cat dander, dog dander, grass pollen mix, and tree pollen mix. BMI at age 3 years was based on study visit measurements.

Metabolomic Profiling

Mass spectrometer platforms, sample extraction and preparation, instrument settings and conditions and data handling were performed at Metabolon, Inc. (Research Triangle Park, NC) and have been described previously in detail(S3). The major components of the process are summarized briefly as follows: The sample preparation process was carried out using an automated MicroLab STAR® system (Hamilton, Reno, NV). Recovery standards were added prior to the first step in the extraction process for quality control purposes. Homogenized fecal samples were methanol extracted on a normalized per mass basis for non-targeted metabolic profiling. Each sample was dried under vacuum to remove the organic solvent. Biochemical levels were measured using ultra-performance liquid chromatography coupled with tandem mass spectrometry (UPLC-MS/MS), as described(S3). The reproducibility of the extraction protocol was assessed by the recovery of the xenobiotic compounds supplemented in every sample prior to extraction. Identification of known chemical entities was based on comparison

to metabolomic library entries of purified standards based on chromatographic properties and mass spectra. Of a total of 737 measured and annotated metabolites, 662 (90%) were annotated based on known standards and the remaining 75 are denoted by an asterisk in relevant results tables.

Additional processing of metabolite data was performed using R version 3.4.0. Metabolites with known chemical identities and interquartile range (IQR) > 0 were analyzed. Intestinal metabolite relative abundances were normalized to sample mass (mg). Missing values were replaced with half of the minimum relative abundance observed for that metabolite. Resulting values were $\log_{10}$ normalized and *pareto*-scaled (mean-centered and divided by the square root of the standard deviation). Principal components analysis of metabolite relative abundances was performed to look for outliers and age trends. One stool sample was over 10 standard deviations from the mean on principal component (PC) 1 and over 6 standard deviations from the mean on PC2 and was excluded from analysis.

Microbiome Profiling

DNA extraction and sequencing of the bacterial 16S V4 hyper-variable region using the Illumina MiSeq platform (San Diego, CA) were performed at Partners Healthcare Personalized Medicine (Boston, MA). Approximately 200 mg of stool per sample was resuspended in 1 mL InhibitEX Buffer from Qiagen QIAamp Fast DNA Stool Mini Kit (Qiagen, Catalogue # 51604) in tubes containing 0.1 mm silicon beads (Genesee Scientific, Catalogue # 31-212S1). Samples were disrupted using a Mini-Beadbeater 24 from Biospec Products for 3 minutes. Extraction

continued on the contents of the entire tube using the Qiagen QIAamp Fast DNA Stool Mini Kit automated on a Qiagen QIAcube. Resulting DNA quality control was evaluated using a PicoGreen assay (Quant-iT dsDNA assay kit, Thermofisher, catalogue number P7589) on a Gemini XP spectrophotometer from Molecular Devices.

Input into library construction was 15 ng DNA using NEXTflex 16S V4 Amplicon-Seq kit 2.0 (Bioo Scientific, Catalogue # 4203-04). Finished libraries were normalized to 10 nM using PicoGreen quantitation and sizing information gathered using a TapeStation D100 screen tapes and reagents (Agilent, Catalogue # 5067-5582 and 5067-5583). Final quality control was carried out to establish the amount of library containing ligated Illumina adaptors using KAPA Library Quantification Kits (KAPA biosystems, Catalogue # KK4824). Up to 288 libraries were pooled to provide equimolar amounts of each library in the pool, which was then run on Illumina MiSeq using MiSeq Reagent Kit v3 600 cycle kit (MS-102-3003). Each pool contained a 20% spike-in of PhiX Control library v3 (Illumina, Catalogue # FC-110-3001) to increase diversity of the library sequence. FASTQ files were generated on the MiSeq instrument.

Filtering, trimming, and chimera checking were performed as previously described(S4, S5). We used closed reference OTU classification in Qiime to group sequences according to sequence similarity and match to taxonomy(S6). Additional processing of microbiome data was performed using Phyloseq package for R(S7). Specifically, samples with read counts of less than 1,000 and OTUs that were not seen at least once in at least 5% of samples were excluded. OTUs with the same species annotation were merged using the tax_glom function.

89

90 As it was recently shown that quantitative microbiome profiling is preferable to relative
91 abundance profiling in co-occurrence analyses and in seeking disease associations(S8),
92 quantitative PCR using universal 16S rRNA primers was performed at Partners Healthcare
93 Personalized Medicine to estimate total bacterial biomass. A standard curve was generated
94 using dilutions of commercially available *E. coli* genomic DNA (Affymetrix/USB, ATCC 11303
95 strain, Santa Clara, CA). Samples with less than 15 ng of total DNA per mL of stool or PCR cycle
96 threshold value less than 8 or greater than 36 were excluded from analysis, and this accounts
97 for why the number of stool samples with microbiome profiling is lower than the number of
98 samples with metabolomic profiling. Species-level taxa relative abundances were converted to
99 fractional abundances, multiplied by estimated bacterial biomass (in ng/microliter of 16S rRNA
100 DNA), multiplied by a constant (10,000) and rounded to the nearest whole number for analysis.

101

102 **Calorie Intake Ascertainment**

103 We used food frequency questionnaire (FFQ) responses to estimate daily caloric intake.
104 Nutrient compositions of FFQ items were determined using the Harvard nutrition composition
105 database, which is based on US Department of Agriculture publications supplemented by other
106 publications and information from manufacturers(S9–S11). Of the 87 items on the FFQ, 4
107 (bacon, margarine tub, margarine stick and cereal (cold)) were not in the Harvard nutrient
108 composition database and nutrient compositions for these foods were obtained from the US
109 Department of Agriculture Nutrient Database for Standard Reference (Release 28)(S11).
110 Standard portion sizes were obtained using the US Department of Agriculture What's In the

Foods You Eat Search Tool(S12). The portion size associated with the label “quantity not specified” was selected for each food item except for pizza, for which 119 grams, the portion size for “1 piece, not further specified,” was selected instead of the “quantity not specified” portion size of 238 grams.

For each of the 87 FFQ items, reported food frequencies were converted to estimated number of daily servings: “Never” was replaced by 0, “< 1 time per week” was replaced with 0.5/7, “Once per week” was replaced with 1/7, “2-4 times/week” was replaced with 3/7, “Nearly daily or daily” was replaced with 1, “2-4 times/day” was replaced with 3, “2 or more times/day” was replaced with 3, and “5 or more times/day” was replaced with 5. Daily intake of calories was calculated by multiplying the estimated number of daily servings for each food item by the calorie content in a standard portion of that food item, and then summing the result across all food items for each subject.

Statistics

Statistical analyses were conducted using R version 3.5.0 (R Foundation for Statistical Computing). Baseline characteristics and potential predictors of the early-life intestinal microenvironment were tabulated and Chi square and *t* tests were used to test for differences in categorical and continuous characteristics, respectively, by phenotype (asthma vs no asthma).

To identify metabolic pathways associated with asthma, we used a network approach to identify highly correlated metabolites, which are expected to be functionally related, using the weighted gene correlation network analysis (WGCNA) R package (version 1.61)(S13). WGCNA identifies modules of highly correlated features in high-dimensional datasets. Signed network construction was performed using Spearman correlation coefficients and applying a minimum module size of 4 metabolites and a soft thresholding power of 12 for metabolite networks (chosen by using the pickSoftThreshold function of the WGCNA R package to achieve a scale-free topology fitting index > 0.9). Eigenvalues summarizing relative abundances of metabolites of each metabolite module for each subject were used in subsequent analyses.

Logistic regression was used to determine the associations of potential predictors of the intestinal microenvironment (such as perinatal antibiotics), individual metabolites, and metabolite module eigenvalues with asthma. Covariates in adjusted analyses were selected on the basis of significant ($p < 0.05$) bivariate associations between these variables and asthma (**Table 1**) and included sex, race/ethnicity, study center, maternal education and gestational age. Maternal education and household income were highly correlated (Chi square p value < 0.01) and both were associated with asthma in bivariate analyses (**Table 1**). Maternal education was included in adjusted models instead of household income, as over 20% of subjects refused to report or did not know their household income (**Table 1**). Sensitivity analyses were performed in which steroid use was added as a covariate to logistic regression models.

Multivariable linear regression analyses were performed to determine associations between breastfeeding and asthma-associated metabolites and metabolite modules. Covariates were selected based on *a priori* knowledge of likely confounders of the breastfeeding-metabolite association and included sex, race/ethnicity, study center, maternal education, gestational age, perinatal antibiotics, mode of delivery, and VDAART trial treatment assignment.

Plasma metabolite summary scores were generated by summing relative abundances for total PUFA (13 metabolites), omega-6 PUFA (6 metabolites), omega-3 PUFA (5 metabolites), endocannabinoids (6 metabolites) and diacylglycerols (19 metabolites). Pearson correlation was used to analyze the association between plasma summary scores and corresponding intestinal metabolite module eigenvalues.

Mediation analyses were performed for two purposes: (1) to estimate the direct association between breastfeeding and asthma, and the indirect association mediated through stool metabolites; and (2) to estimate the direct association between dietary variables and asthma, and the indirect association mediated through intestinal metabolites. Analyses were adjusted for sex, race/ethnicity, study center, maternal education and gestational age. Dietary variable score was log-normalized. Mediation analysis assumes no unmeasured confounding or effect modification between modeled variables. The R package “mediation” was used and 95% confidence intervals were based on quasi-Bayesian approximation with 2,000 Monte Carlo draws(S14).

All tests were 2-sided and the significance level was pre-specified at $p < 0.05$. Given the exploratory nature of this analysis, adjustments were not made for multiple comparisons unless otherwise specified.

Integrative Analyses

Coinertia analysis was performed using the ade4 R package(S15) to compare global associations of the plasma metabolome, gut microbiome and diet with the intestinal metabolome. This method identifies successive axes of covariance between two sets of data available for a single group of subjects. It generates an RV score, which ranges from 0 to 1 with higher scores indicating greater global similarity between data sets. The RV.rtest function was used to test for significance of the RV score based on 99 Monte-Carlo simulations. Because the RV score increases with decreasing sample size(S16), analysis was restricted to subjects with all data types: intestinal and plasma metabolome, intestinal microbiome and diet ($n = 178$).

Sparse canonical correlation analysis was performed using the PMA R package(S17) to identify features of the intestinal metabolome that are correlated both with features of other data types (plasma metabolome, intestinal microbiome, diet) and with asthma. This method estimates the linear relationship between two sets of data from a single group of subjects. For each pair of data types, a limited number of features were assigned non-zero canonical loadings on the first canonical variate, which were used to calculate canonical scores for each subject for each data type. Pearson correlation was used to ensure that canonical scores were indeed correlated between each pair of data types, and the association of each canonical score with

197 asthma was queried in adjusted logistic regression analyses. Spearman correlation heatmaps of
198 features with non-zero canonical loadings were constructed using R function heatmap.

199

References:

- S1. Litonjua AA, Carey VJ, Laranjo N, Harshfield BJ, McElrath TF, O'Connor GT, et al. Effect of Prenatal Supplementation With Vitamin D on Asthma or Recurrent Wheezing in Offspring by Age 3 Years: The VDAART Randomized Clinical Trial. *JAMA* 2016;315:362–70.
- S2. Litonjua AA, Lange NE, Carey VJ, Brown S, Laranjo N, Harshfield BJ, et al. The Vitamin D Antenatal Asthma Reduction Trial (VDAART): Rationale, design, and methods of a randomized, controlled trial of vitamin D supplementation in pregnancy for the primary prevention of asthma and allergies in children. *Contemp Clin trials* 2014;38:37–50.
- S3. Evans AM, Bridgewater BR, Liu Q, Mitchell MW, Robinson RJ, Dai H, et al. Metabolomics: Open access high resolution mass spectrometry improves data quantity and quality as compared to unit mass resolution mass spectrometry in high-throughput profiling metabolomics. *Metabolomics* 2014;4:1–7.
- S4. Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* 2012;486:215–21.
- S5. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* 2011;21:494–504.
- S6. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7:335–6.
- S7. McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 2013;8(4):e61217.

- 222 S8. Vandeputte D, Kathagen G, D'Hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, et al.
 223 Quantitative microbiome profiling links gut community variation to microbial load.
 224 *Nature* 2017;551:507–511.
- 225 S9. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, et al. Reproducibility
 226 and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol*
 227 1985;122:51–65.
- 228 S10. Oken E, Kleinman KP, Olsen SF, Rich-Edwards JW, Gillman MW. Associations of seafood
 229 and elongated n-3 fatty acid intake with fetal growth and length of gestation: Results
 230 from a US pregnancy cohort. *Am J Epidemiol* 2004;160:774–783.
- 231 S11. USDA. USDA National Nutrient Database for Standard Reference Release 28.
 232 <https://ndb.nal.usda.gov/ndb/>. May, 2016. Accessed January 2, 2018.
- 233 S12. USDA. What's In The Foods You Eat Search Tool, 2013-2014.
 234 <https://reedir.arsnet.usda.gov/codesearchwebapp>. Accessed January 2, 2018.
- 235 S13. Langfelder P, Horvath S, Fisher R, Zhou X, Kao M, Wong W, et al. WGCNA: an R package
 236 for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
- 237 S14. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R package for causal
 238 mediation analysis. *J Stat Softw* 2014;59:1–38.
- 239 S15. Dray S, Dufour AB. The ade4 Package: Implementing the duality diagram for ecologists. *J*
 240 *Stat Softw* 2007;22(4):1-20.
- 241 S16. Fruciano C, Franchini P, Meyer A. Resampling-based approaches to study variation in
 242 morphological modularity. *PLoS One* 2013;8(7):e69376.
- 243 S17. Witten DM, Tibshirani R, Hastie T. A penalized matrix decomposition, with applications to

244 sparse principal components and canonical correlation analysis. *Biostatistics*
245 2009;10:515–534.