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.

					Including Steroid Use in Model	
Metabolite	**Pathway**	**Subpathway**	**Odds Ratio (95% Confidence Interval)**	***p* value**	**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-glycolylneuraminate	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	hexadecasphinganine (d16:0)*	valerylphenylalanine
linoleoyl ethanolamide	ceramide (d18:1/14:0, d16:1/16:0)*	N-acetylphenylalanine
stearoyl ethanolamide	heptadecasphingosine (d17:1)	3-phenylpropionate (hydrocinnamate)
palmitoyl ethanolamide	N-palmitoyl-sphingosine (d18:1/16:0)	indole
arachidoyl ethanolamide (20:0)*	N-butyroyl-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	hexadecasphingosine (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-methylpipecolate	**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-acetylornithine
3-methylhistidine	N-formylmethionine	2-methylserine
N-acetyl-3-methylhistidine*	N-alpha-acetylornithine	alanine
anserine	histidine	**Module 27**
Module 24	cysteine	glycodeoxycholate
gluconate	indoleacetate	linolenoylcarnitine (C18:3)*
tricarballylate	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	glycyrrhetinate	3,7-dimethylurate
alpha-CEHC	LAHSA (18:2/OH-18:0)*	7-methylurate
phytosphingosine	OAHSA (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-ketoetiocholanolone sulfate	trans-nonadecenoate (tr 19:1)*
3-hydroxyphenylacetate	cholate sulfate	10-hydroxystearate
lactobionate	pregnanolone/allopregnanolone sulfate	15-methylpalmitate
2-(4-hydroxyphenyl)propionate	I-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	isovalerylglycine	nonadecanoate (19:0)
nicotinate	2-methylbutyrylglycine	margarate (17:0)
3-deoxyoctulosonate	isobutyrylglycine	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	tryptophylglycine	**Module 41**
pantothenate	lysylleucine	urate
arabitol/xylitol	uridine-2',3'-cyclic monophosphate	S-1-pyrroline-5-carboxylate
pyruvate	1-methylnicotinamide	guanidinosuccinate
mannitol/sorbitol	leucylglycine	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
cysteinylglycine	valylleucine	3-ketosphinganine
3-hydroxybutyrate (BHBA)	isoleucylglycine	4-imidazoleacetate
N6,N6,N6-trimethyllysine	valylglutamine	alpha-ketobutyrate
6-hydroxynicotinate	phenylalanylalanine	N-methyltryptophan
3,4-dihydroxyphenylacetate	phenylalanylglycine	3-methyl-2-oxobutyrate
N-acetylhistamine	leucylalanine	4-hydroxyphenylpyruvate
3-hydroxyvalerate	valylglycine	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-methylglutarylcarnitine (2)	N-(2-furoyl)glycine	sucralose
succinylcarnitine (C4-DC)	creatine	gamma-tocopherol/beta-tocopherol
linoleoylcarnitine (C18:2)*	hippurate	saccharopine
carnitine	dimethylglycine	harmane
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	carnosine	2-hydroxypalmitate
succinate	creatinine	dihydroorotate
alpha-hydroxyisocaproate	benzoate	pentadecanoate (15:0)
3-(4-hydroxyphenyl)lactate	1-methylhistidine	4-cholesten-3-one
phenylacetylglycine	**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.

	Association with Breastfeeding		Association with Asthma		Mediation Analysis Results*	
Metabolite	**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%
17α-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).

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

	Canonical Loading	
Food	**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.

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

Calorie Intake Ascertainment

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

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

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 auantity and auality 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.

S8. Vandeputte D, Kathagen G, D'Hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, et al. Quantitative microbiome profiling links gut community variation to microbial load. *Nature* 2017;551:507–511.

S9. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985;122:51–65.

S10. Oken E, Kleinman KP, Olsen SF, Rich-Edwards JW, Gillman MW. Associations of seafood and elongated n-3 fatty acid intake with fetal growth and length of gestation: Results from a US pregnancy cohort. *Am J Epidemiol* 2004;160:774–783.

S11. USDA. USDA National Nutrient Database for Standard Reference Release 28. https://ndb.nal.usda.gov/ndb/. May, 2016. Accessed January 2, 2018.

S12. USDA. What's In The Foods You Eat Search Tool, 2013-2014. https://reedir.arsnet.usda.gov/codesearchwebapp. Accessed January 2, 2018.

S13. Langfelder P, Horvath S, Fisher R, Zhou X, Kao M, Wong W, et al. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.

S14. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R package for causal mediation analysis. *J Stat Softw* 2014;59:1–38.

S15. Dray S, Dufour AB. The ade4 Package: Implementing the duality diagram for ecologists. *J Stat Softw* 2007;22(4):1-20.

S16. Fruciano C, Franchini P, Meyer A. Resampling-based approaches to study variation in morphological modularity. *PLoS One* 2013;8(7):e69376.

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.