nature medicine

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41591-018-0216-2

In the format provided by the authors and unedited.

Infant diet and maternal gestational weight gain predict early metabolic maturation of gut microbiomes

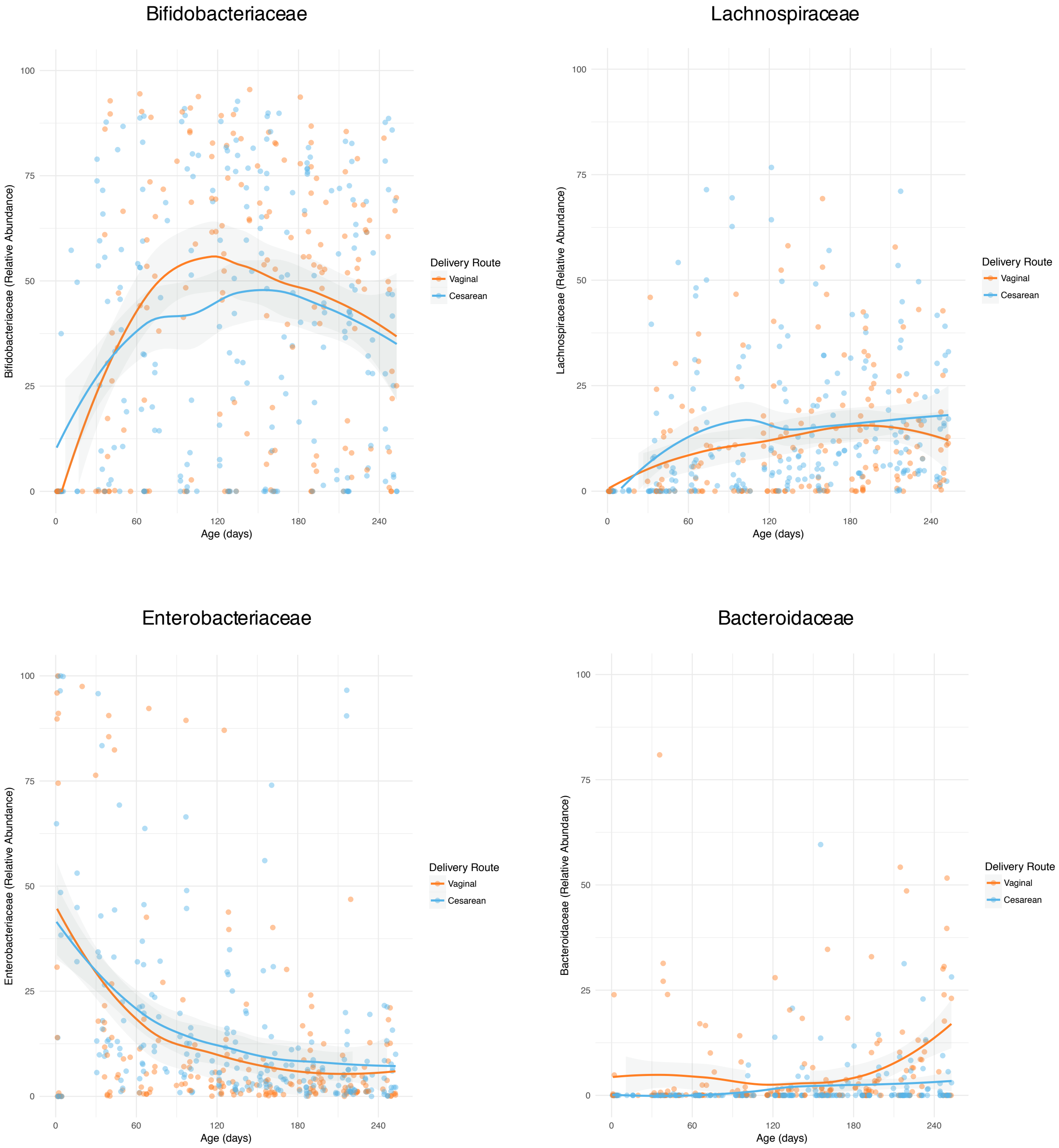
Aimee M. Baumann-Dudenhoeffer[1,2*], Alaric W. D'Souza[2], Phillip I. Tarr[3,4], Barbara B. Warner[1] and Gautam Dantas[2,4,5,6*]

[1]Division of Newborn Medicine, Department of Pediatrics, Washington University in St. Louis School of Medicine, St. Louis, MO, USA. [2]The Edison Family Center for Genome Sciences and Systems Biology, Washington University in St. Louis School of Medicine, St. Louis, MO, USA. [3]Division of Gastroenterology, Hepatology, and Nutrition, Department of Pediatrics, Washington University in St. Louis School of Medicine, St. Louis, MO, USA. [4]Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, MO, USA. [5]Department of Pathology and Immunology, Washington University in St. Louis School of Medicine, St. Louis, MO, USA. [6]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, USA. *e-mail: dudenhoeffer@wustl.edu; dantas@wustl.edu

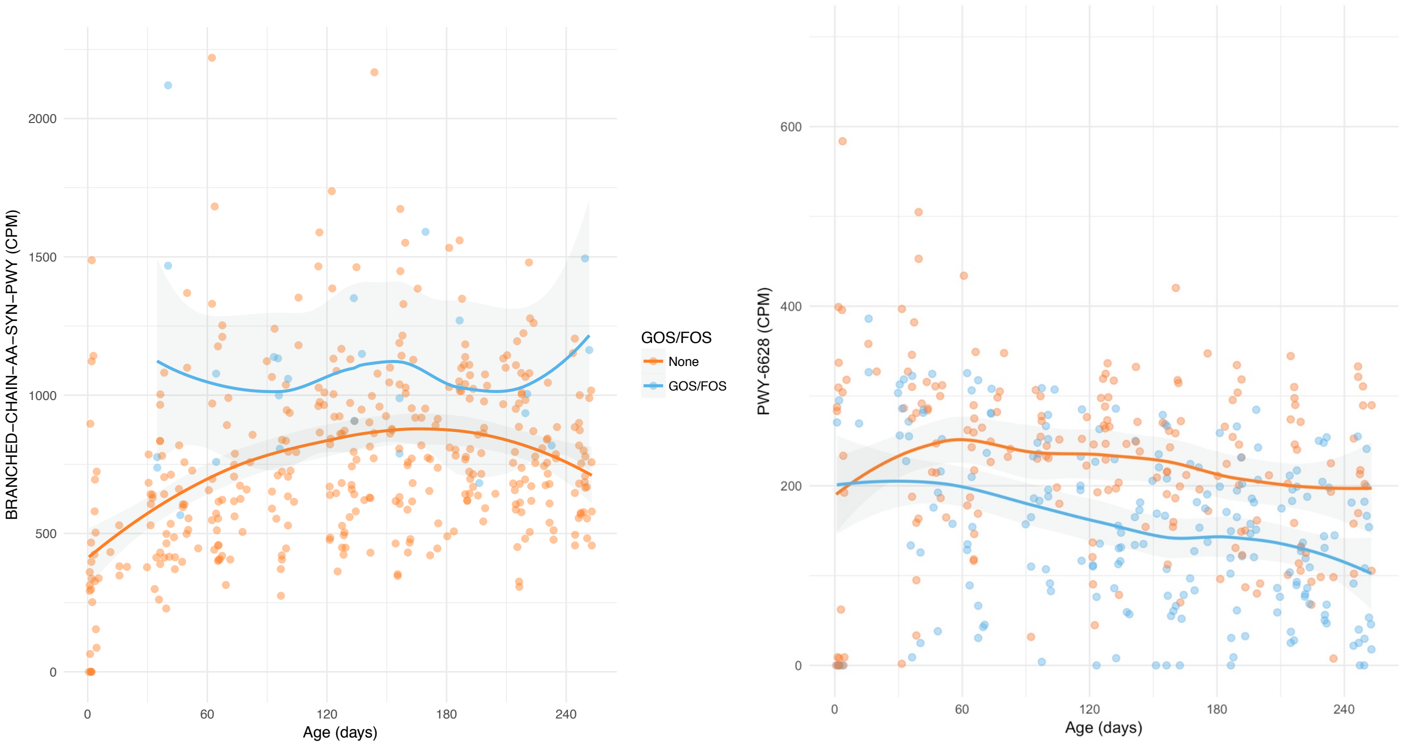
Infant Diet and Maternal Gestational Weight Gain Predict Early Metabolic Maturation of Gut Microbiomes

Aimee M. Baumann-Dudenhoeffer, Alaric W. D’Souza, Phillip I. Tarr , Barbara B. Warner, Gautam Dantas

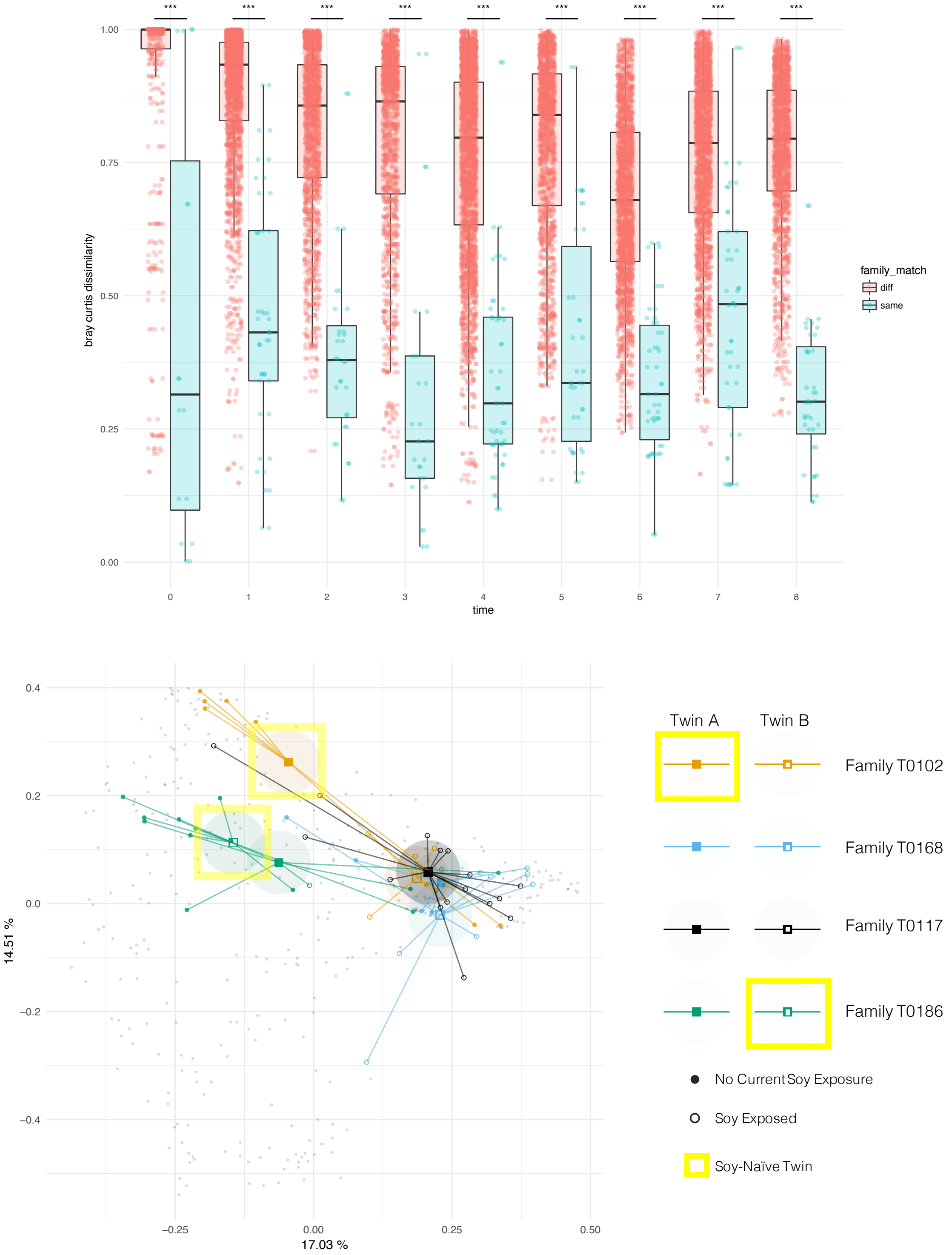
Supplementary Information



Supplementary Figure 1: Relative Abundance of Major Taxa by Age and Route of Delivery
The y-axis represents the relative abundance of major taxonomic families, while the x-axis represents age in days. Regression lines with 95% confidence intervals (light gray) are drawn using the loess method in R. All panels compare vaginally delivered infants (N=175 samples, orange) with infants born via Cesarean section (N=227, blue). Bifidobacteriaceae are shown top left, Lachnospiraceae top right, Enterobacteriaceae bottom left, and Bacteroidaceae bottom right. Only Bacteroidaceae had a statistically significant negative correlation with Cesarean delivery (p=0.003). All p values are two-tailed, from multivariate maximum-likelihood GLMMs, Tukey-corrected for multiple comparisons.

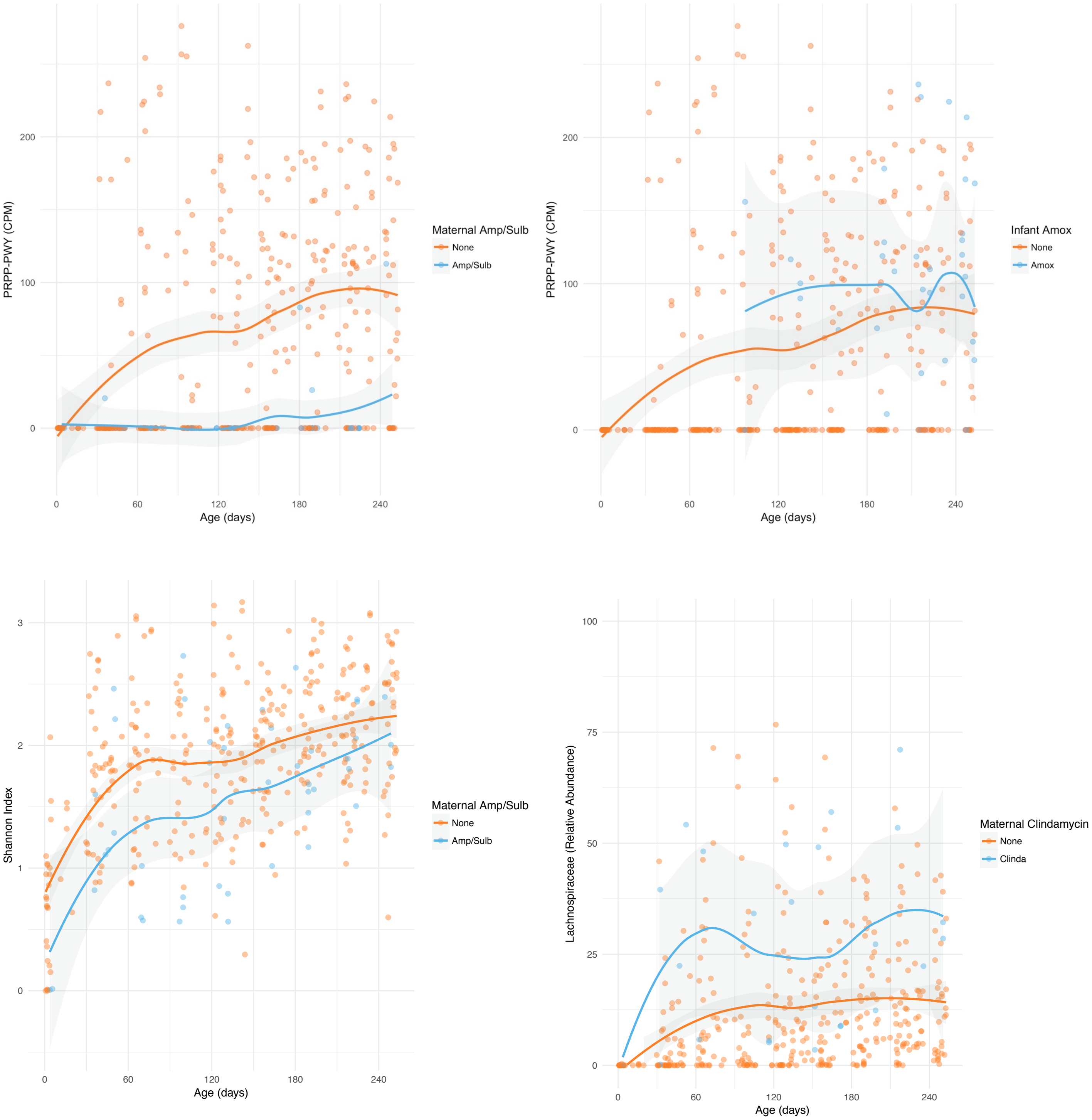


Supplementary Figure 2: Selected Functional Pathways by Age and Prebiotic Exposure
The y-axis represents the abundance of selected pathways in normalized cpm (counts per million), while the x axis represents age in days. Regression lines with 95% confidence intervals (light gray) are drawn using the loess method in R. The left panel plots BCAA synthesis pathway (BRANCHED-CHAIN-AA-SYN-PWY) abundance for currently GOS/FOS exposed infants (blue, N=26) and non-exposed infants (orange, N=376); there is a significant positive association with GOS/FOS exposure ($p<0.001$). The right shows phenylalanine synthesis pathway (PWY-6628) abundance for infants with any lifetime GOS exposure (blue, N=204) and lifetime GOS-naïve infants (orange, N=198). There was a significant negative association between PWY-6628 and an interaction term between lifetime GOS exposure and time (GOS:log(Day of Life) $p=0.004$); the negative correlation became more significant with time. GOS exposure alone did not have a statistically significant correlation with PWY-6628. All p values are two-tailed, from multivariate longitudinal maximum-likelihood GLMMs, Tukey-corrected for multiple comparisons.

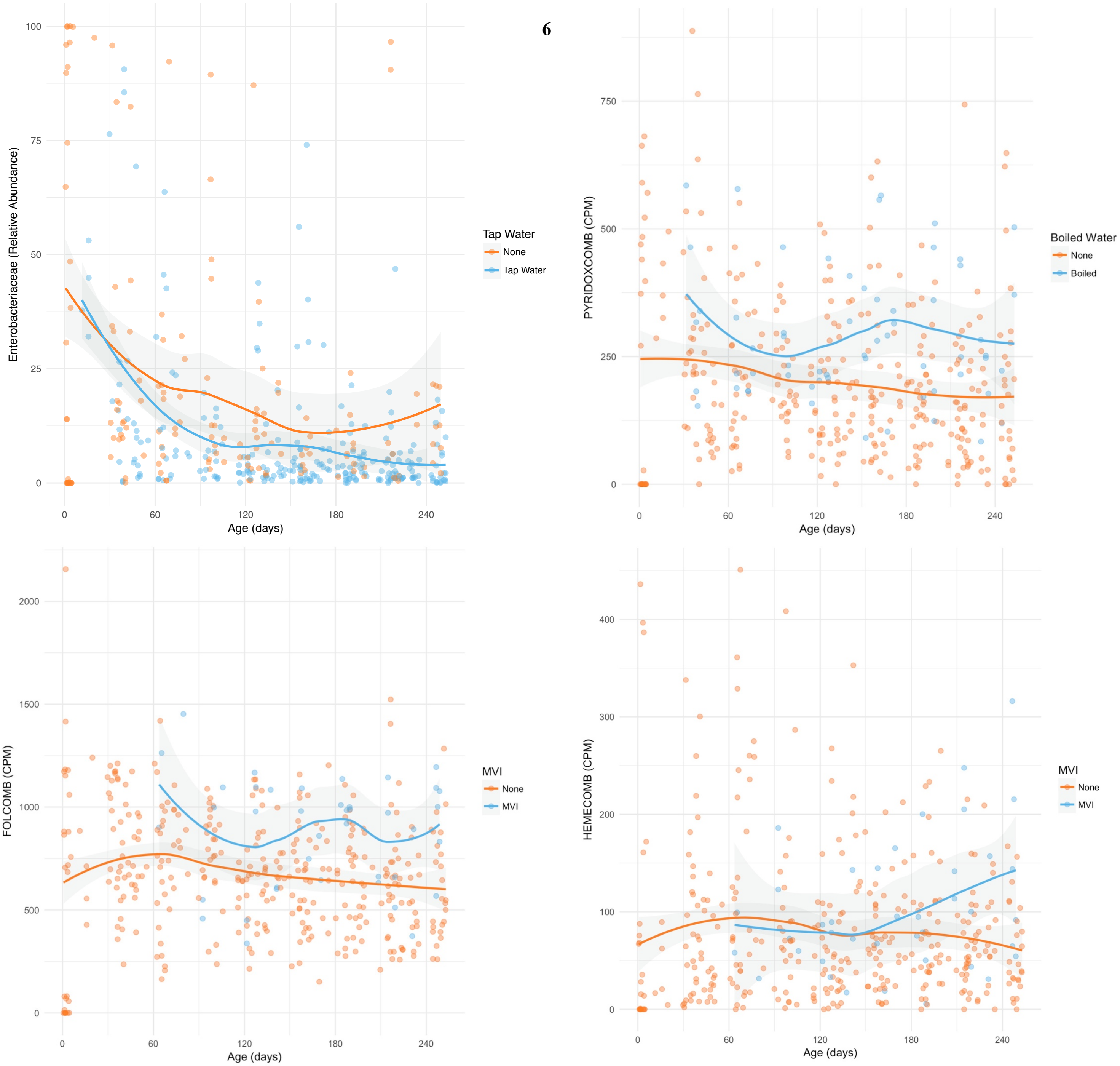


Supplementary Figure 3: Similarity of Twin Microbiomes and Clustering of Soy-Exposed Infants
The top panel compares the taxonomic similarity of twin siblings' microbiomes to the similarity between unrelated infants of approximately the same age. The y-axis represents the Bray-Curtis dissimilarity index (calculated based on species-level community composition; the x axis represents age in months. The Bray-Curtis dissimilarity index is shaded blue for sibling pairs and orange for unrelated age-matched infants. At all ages, infants' microbiomes are significantly more similar to those of their twin sibling than to those of unrelated infants (p<0.001 at each month of life, two-tailed Welch two-sample t-test). See https://bitbucket.org/alaricwdsouza/twindiet/src/master/TwinDiet_compareTwins.R

The bottom panel is a Principal Coordinate Analysis (PCoA) plot based on a species-level Bray-Curtis dissimilarity matrix, colored to show the four soy-exposed families. For each subject, a centroid for all timepoints is plotted with either a solid square (Twin A) or hollow square (Twin B), and segments are drawn connecting the centroid to each sample from that subject. Soy-naïve siblings from soy-discordant families are highlighted with a yellow box. Soy-exposed infants cluster together on the right side of the plot; the soy-exposed Twin B from family T0102 qualitatively resembles unrelated soy-exposed infants more than their soy-naïve twin sibling. Twin A from Family T0186 was only soy-exposed at a single timepoint, and recovered to resemble their soy-naïve twin.

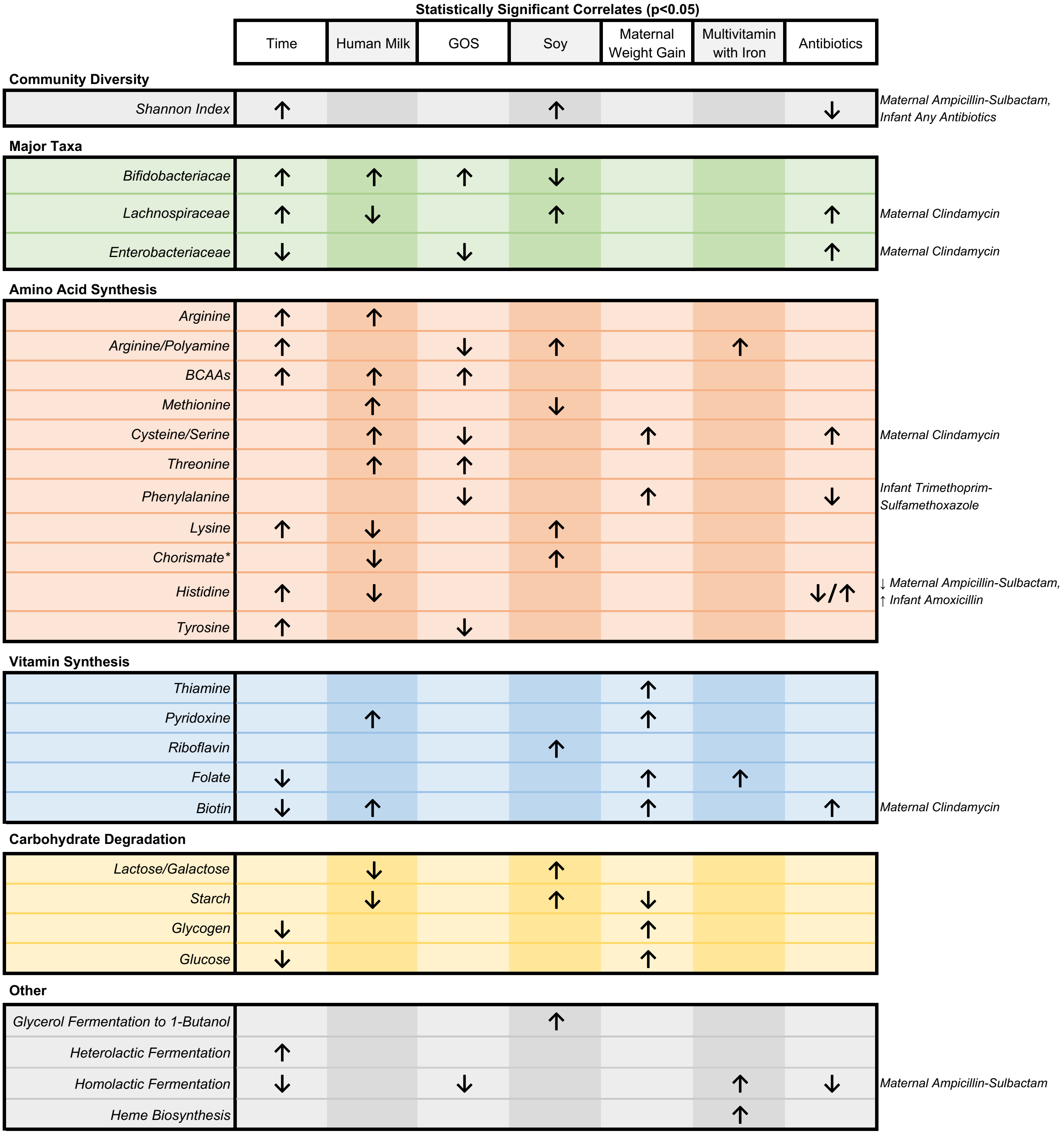


Supplementary Figure 4: Selected Microbial Community Features by Age and Antibiotic Exposure
The y axis represents either relative abundance of major taxa, alpha diversity (Shannon index), or pathway abundance in normalized counts per million. The x-axis represents age in days. Regression lines with 95% confidence intervals (light gray) are drawn with the loess method in R. The top two panels show histidine-purine-pyrimidine synthesis pathway (PRPP-PWY) abundance colored by pre- and post-natal antibiotic exposures. The top left panel contrasts infants whose mothers received intrapartum ampicillin-sulbactam (blue, N=46) with those whose mothers did not (orange, N=356); PRPP-PWY was negatively correlated with maternal ampicillin-sulbactam exposure (p=0.012). The top right panel shows infants postnatally exposed to amoxicillin (blue, N=38) and amoxicillin-naïve infants (orange, N=364); PRPP-PWY had a significant positive association with amoxicillin exposure (0=0.011). The bottom left panel shows the Shannon index, colored by maternal intrapartum ampicillin-sulbactam-exposed (blue) and naïve (orange) status; species diversity was negatively correlated with ampicillin-sulbactam exposure (p=0.005). The bottom right panel shows relative abundance of Lachnospiraceae, colored by maternal intrapartum clindamycin exposed (blue, N=25) and naïve status (orange, N=377); Lachnospiraceae were positively correlated with intrapartum clindamycin (p=0.008). All p values are two-tailed, from multivariate longitudinal maximum-likelihood GLMMs, Tukey-corrected for multiple comparisons.



Supplementary Figure 5: Selected Taxa and Pathways by Age, Water, and Multivitamin with Iron (MVI) Exposure
The y axis represents the relative abundance of selected taxa or abundance of pathways in normalized cpm (counts per million). The x-axis represents age in days. Regression lines with 95% confidence intervals (light gray) are drawn using the loess method in R. All p values are two-tailed from longitudinal multivariate maximum-likelihood GLMMs, Tukey-corrected for multiple comparisons. The top left panel shows Enterobacteriaceae colored by lifetime tap water-exposed (blue, N=251) and tap water-naïve status (orange, N=151). Enterobacteriaceae negatively correlated with an interaction term between tap water exposure and time (Tap Water:log(Day of Life) p=0.016); the association is more negative with increasing age. The top right panel shows the aggregate pyridoxine synthesis variable PYRIDOXCOMB colored by lifetime boiled/distilled water exposed (blue, N=61) and boiled/distilled water-naïve status (orange, N=341); PYRIDOXCOMB is significantly positively associated with boiled/distilled water exposure (p=0.003).

In both bottom plots, infants exposed to multivitamin with iron (MVI) are shaded blue (N=40), all others are shaded orange. (N=362). The bottom left panel shows a significant positive correlation between MVI exposure and the aggregate folate synthesis pathway FOLCOMB (p<0.001). The right panel shows a positive correlation of the aggregate heme synthesis pathway variable HEMECOMB and an interaction term between MVI exposure and time (MVI:log(Day of Life) p=0.026); the association became more positive with increasing age. MVI alone did not have a significant association with HEMECOMB.



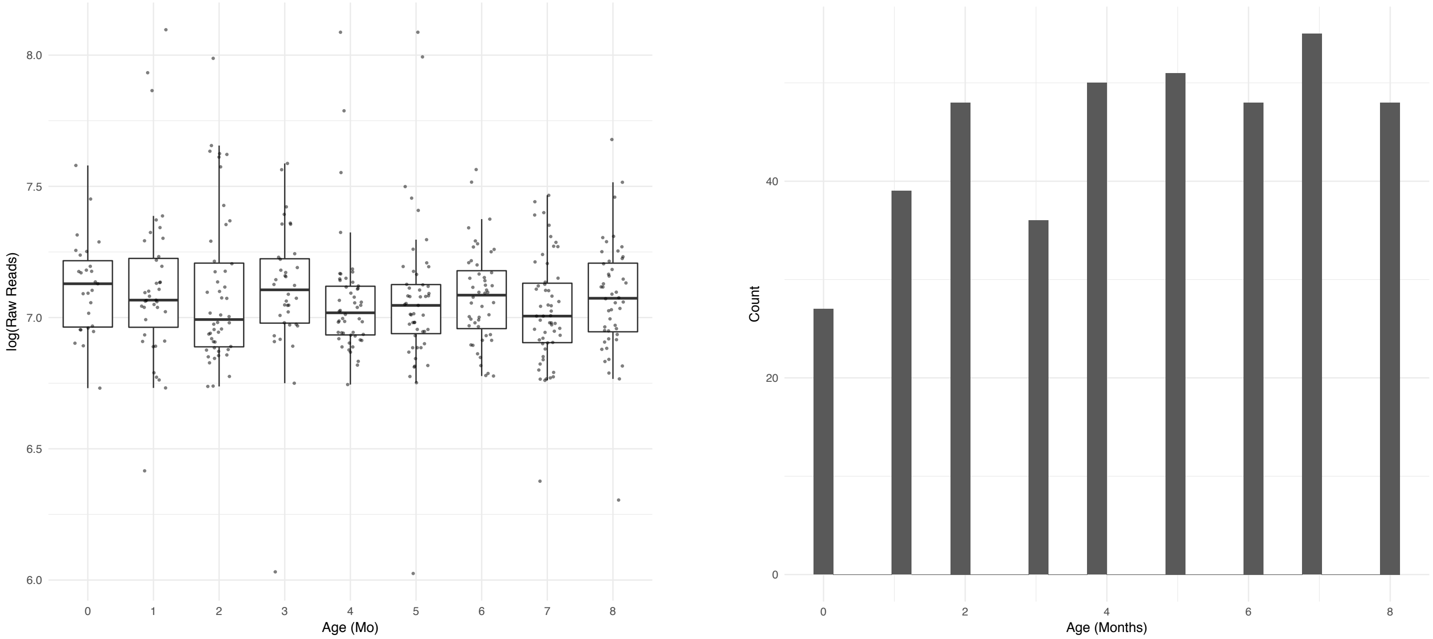
Statistically Significant Correlates (p<0.05)

	Time	Human Milk	GOS	Soy	Maternal Weight Gain	Multivitamin with Iron	Antibiotics	
Community Diversity								
Shannon Index	↑			↑			↓	*Maternal Ampicillin-Sulbactam, Infant Any Antibiotics*
Major Taxa								
Bifidobacteriacae	↑	↑	↑	↓				
Lachnospiraceae	↑	↓		↑			↑	*Maternal Clindamycin*
Enterobacteriaceae	↓		↓				↑	*Maternal Clindamycin*
Amino Acid Synthesis								
Arginine	↑	↑						
Arginine/Polyamine	↑		↓	↑		↑		
BCAAs	↑	↑	↑					
Methionine		↑		↓				
Cysteine/Serine		↑	↓		↑		↑	*Maternal Clindamycin*
Threonine		↑	↑					
Phenylalanine			↓		↑		↓	*Infant Trimethoprim-Sulfamethoxazole*
Lysine	↑	↓		↑				
*Chorismate**		↓		↑				
Histidine	↑	↓					↓/↑	↓ *Maternal Ampicillin-Sulbactam,* ↑ *Infant Amoxicillin*
Tyrosine	↑		↓					
Vitamin Synthesis								
Thiamine					↑			
Pyridoxine		↑			↑			
Riboflavin				↑				
Folate	↓				↑	↑		
Biotin	↓	↑			↑		↑	*Maternal Clindamycin*
Carbohydrate Degradation								
Lactose/Galactose		↓		↑				
Starch		↓		↑	↓			
Glycogen	↓				↑			
Glucose	↓				↑			
Other								
Glycerol Fermentation to 1-Butanol				↑				
Heterolactic Fermentation	↑							
Homolactic Fermentation	↓		↓			↑	↓	*Maternal Ampicillin-Sulbactam*
Heme Biosynthesis						↑		

Supplementary Figure 6: Summary of Observed Significant Pre- and Post-Natal Predictors of Infant Gut Microbiome Maturation

This image qualitatively summarizes statistically significant associations ($p<0.05$) between major infant gut microbiome taxa and gene-encoded functional pathways with clinical variables. All p values are two-tailed, from longitudinal multivariate maximum-likelihood GLMMs, Tukey-corrected for multiple comparisons. Upward arrows indicate a positive association; downward arrows indicate a negative association. Maternal antibiotics refer to intrapartum antibiotic exposures. Exact p values, coefficients, confidence intervals, and sample sizes are found in Tables S3 and S5.

*Chorismate is not an amino acid, but is a precursor to tryptophan, tyrosine, and phenylalanine



Supplementary Figure 7: Sample Sequencing Characteristics, by Month of Life
The left panel shows $\log_{10}$(raw reads) for all 402 samples, distributed by month of life. There was no age-related bias in raw sampling depth. The right panel shows a histogram of the number of samples included in the study, by month of life.

Supplementary Table 1: Clinical Characteristics of Study Population

N= number of infants with a given characteristic. This differs from the main text and other Figures and Tables, where N refers to the number of samples with a given characteristic.

1A: General Demographic Characteristics (Median, IQR, Range Min-Max)

	MEDIAN	IQR	RANGE
Maternal age (y)	29.5	5	20-37
Gestational age (wks)	37	1	34-38
Birthweight (g)	2735	571	1819-3373
Birth length (cm)	49	2.25	45-53
Birth OFC (cm)	33	2	30-45

1B: Number of Subjects Exposed to Maternal Antibiotics

MATERNAL INTRAPARTUM ANTIBIOTICS	**ANY ANTIBIOTICS**	N=46 (77%)
	CEFAZOLIN	N=24 (40%)
	AMPICILLIN	N=6 (10%)
	PENICILLIN G	N=6 (10%)
	VANCOMYCIN	N=2 (3%)
	CLINDAMYCIN	N=4 (7%)
	AMPICILLIN-SULBACTAM	N=8 (13%)

1C: General Demographic Characteristics (Number of subjects, Percent of subjects)

DELIVERY ROUTE		ZYGOSITY*		DISPOSITION (Nursery)		FEEDING ROUTE (Nursery)		SEX		ETHNICITY		RACE	
Vaginal	N=26 (43%)	**Monozygotic**	N=28 (47%)	**Nursery**	N=52 (87%)	**Oral**	N=57 (95%)	**Male**	N=29 (48%)	**Hispanic**	N=2 (3%)	**Black**	N=10 (17%)
Cesarean	N=34 (57%)	**Dizygotic**	N=30 (50%)	**Special Care**	N=8 (13%)	**Oral/gavage**	N=3 (5%)	**Female**	N=31 (52%)	**Non-Hispanic**	N=58 (97%)	**White**	N=50 (83%)

**Two infants had unknown zygosity*

1D: Clinical Characteristics, by Maternal Gestational Weight Gain Quartile

GESTATIONAL WEIGHT GAIN QUARTILE	N	GESTATIONAL WEIGHT GAIN RANGE (KG)	MATERNAL PRE-PREGNANCY BMI	ESTIMATED GESTATIONAL AGE AT DELIVERY (WKS)	MATERNAL AGE (YEARS)	INFANT BIRTHWEIGHT (G)
			Mean +/- SD	*Mean +/- SD*	*Mean +/- SD*	*Mean +/- SD*
1	14	2-15	28.40 +/- 5.31	36.43 +/- 1.22	30.57 +/- 4.73	2670.50 +/- 319.38
2	14	16-20	24.03 +/- 3.17	36.76 +/- 0.75	32.29 +/- 2.97	2701.29 +/- 277.65
3	16	21-24	24.79 +/- 3.05	37.00 +/- 0.73	29.50 +/- 2.68	2687.75 +/- 367.60
4	16	26-33	27.71 +/- 10.08	36.88 +/- 0.73	28.69 +/- 3.77	2800.81 +/- 393.26

GESTATIONAL WEIGHT GAIN QUARTILE	N	GESTATIONAL WEIGHT GAIN RANGE (KG)	MATERNAL DIABETES	MATERNAL PREECLAMPSIA	MONO-ZYGOTIC	DI-ZYGOTIC	UNKNOWN ZYGOSITY	CESAREAN DELIVERY	VAGINAL DELIVERY
			N (%)	*N (%)*	*N (%)*	*N (%)*	*N (%)*	*N (%)*	*N (%)*
1	14	2-15	0 (0.00%)	2 (14.29%)	8 (57.14%)	6 (42.86%)		5 (35.71%)	9 (64.29%)
2	14	16-20	2 (14.29%)	0 (0.00%)	4 (28.57%)	8 (57.14%)	2 (14.29%)	8 (57.14%)	6 (42.86%)
3	16	21-24	2 (12.50%)	4 (25.00%)	10 (62.5%)	6 (37.50%)		14 (87.50%)	2 (12.50%)
4	16	26-33	0 (0.00%)	0 (0.00%)	6 (37.50%)	10 (62.50%)		7 (43.75%)	9 (56.25%)

Supplementary Table 2: Component Pathways of Aggregate Metabolic Function Variables

AGGREGATE PATHWAY	COMPONENT PATHWAYS
Aggregate L-Arginine Synthesis (ARGCOMB)	**ARGSYN-PWY: L-arginine biosynthesis I (via L-ornithine)**
	ARGSYNBSUB-PWY: L-arginine biosynthesis II (acetyl cycle)
	PWY-5154: L-arginine biosynthesis III (via N-acetyl-L-citrulline)
	PWY-7400: L-arginine biosynthesis IV (archaebacteria)
Aggregate Arginine/Polyamine Synthesis (ARGPOLYCOMB)	**ARG+POLYAMINE-SYN: superpathway of arginine and polyamine biosynthesis**
	POLYAMSYN-PWY: superpathway of polyamine biosynthesis I
	POLYAMINSYN3-PWY: superpathway of polyamine biosynthesis II
Aggregate L-Isoleucine Synthesis (ISOLEUCCOMB)	**ILEUSYN-PWY: L-isoleucine biosynthesis I (from threonine)**
	PWY-3001: superpathway of L-isoleucine biosynthesis I
	PWY-5101 L-isoleucine biosynthesis II
	PWY-5103: L-isoleucine biosynthesis III
	PWY-5104 L-isoleucine biosynthesis IV
Aggregate L-Lysine Synthesis (LYSCOMB)	**DAPLYSINESYN-PWY: L-lysine biosynthesis I**
	PWY-2941: L-lysine biosynthesis II
	PWY-2942: L-lysine biosynthesis III
	PWY-5097: L-lysine biosynthesis VI
Aggregate L-Methionine Synthesis (METCOMB)	**HOMOSER-METSYN-PWY: L-methionine biosynthesis I**
	HSERMETANA-PWY: L-methionine biosynthesis III
	METSYN-PWY: L-methionine biosynthesis I
	PWY-5347: superpathway of L-methionine biosynthesis (transsulfuration)
	PWY-5345: superpathway of L-methionine biosynthesis (by sulfhydrylation)
Aggregate S-Adenosyl L-Methionine Synthesis (SAM1COMB)	**PWY-6151: S-adenosyl-L-methionine cycle I**
	MET-SAM-PWY: superpathway of S-adenosyl-L-methionine biosynthesis
Aggregate Pyridoxine Synthesis (PYRIDOXCOMB)	**PYRIDOXSYN-PWY: pyridoxal 5'-phosphate biosynthesis I**
	PWY0-845: superpathway of pyridoxal 5'-phosphate biosynthesis and salvage
Aggregate Thiamine Synthesis (THISYNCOMB)	**THISYN-PWY: superpathway of thiamin diphosphate biosynthesis I**
	PWY-7282: 4-amino-2-methyl-5-phosphomethylpyrimidine biosynthesis
	PWY-6897: thiamine salvage II
	PWY-6892: thiazole biosynthesis
Aggregate Biotin Synthesis (BIOTINCOMB)	**BIOTIN-BIOSYNTHESIS-PWY: superpathway biotin biosynthesis I**
	PWY-6519: 8-amino-7-oxononanoate biosynthesis I
Aggregate Folate Synthesis (FOLCOMB)	**1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis**
	FOLSYN-PWY: superpathway of tetrahydrofolate biosynthesis and salvage
	PWY-6147: 6-hydroxymethyl-dihydropterin diphosphate biosynthesis I
	PWY-6612: superpathway of tetrahydrofolate biosynthesis
Aggregate Lactose/Galactose Degradation (LACGALACCOMB)	**PWY66-422: D-galactose degradation V (Leloir pathway)**
	PWY-6317: galactose degradation I (Leloir pathway)
	LACTOSECAT-PWY: lactose and galactose degradation I
Aggregate Starch Degradation (STARCHCOMB)	**PWY-6737: starch degradation V**
	PWY-6731: starch degradation III
Aggregate Glucose Degradation (GLUCCOMB)	**GLUCOSE1PMETAB-PWY: glucose and glucose-1-phosphate degradation**
	PWY-6901: superpathway of glucose and xylose degradation
Aggregate Sucrose Degradation (SUCCOMB)	**PWY-621: sucrose degradation III (sucrose invertase)**
	PWY-5384: sucrose degradation IV (sucrose phosphorylase)
Aggregate Heme Synthesis (HEMECOMB)	**HEME-BIOSYNTHESIS-II: heme biosynthesis I (aerobic)**
	HEMESYN2-PWY: heme biosynthesis II (anaerobic)
	PWY-5918: superpathay of heme biosynthesis from glutamate
	PWY-5920: superpathway of heme biosynthesis from glycine
	PWY0-1415: superpathway of heme biosynthesis from uroporphyrinogen-III

Supplementary Table 3: Maximum-Likelihood Longitudinal Multivariate GLMM Model Information
This table is too large to be displayed, and is located in an appended Excel file. Pseudo-R^2 (calculated using the r.squaredGLMM() function in the R MuMin package) scaled residuals, sample size, random effects, and fixed effects are listed for all longitudinal GLMMs in this manuscript. Parameters coefficients, confidence intervals, unadjusted two-tailed p-values, and p-values Tukey-corrected for multiple comparisons are shown for all fixed effects. Residuals, random effects, and unadjusted fixed effects are generated using the lmer() function in the R lme4 package. Tukey correction for multiple comparisons is done using the glht() function in the R multcomp package.

Supplementary Table 4: Pathway-Identified Taxa
This table is too large to be displayed, and is located in an appended Excel file. All taxa and taxa-identified pathways are inferred using HUMAnN2, as described in the Online Methods. Column 4A shows proportional representation of individual subpathways within aggregate variables, based on normalized counts per million for individual pathways and aggregate variables. Percentages express proportion of total counts associated with an individual pathway within an aggregate variable. Counts not identified with any taxon are included. Column 4B shows proportional contributions of genera homologous to individual metabolic pathways, based on normalized counts per million for individual pathways. Percentages express proportion of total counts associated with a genus, and do not include counts not identified with any genera. Genera comprising >=1% of taxa-IDed pathways are included. Column 4C shows pathways with no homologous taxon identified.

Supplementary Table 5: Qualitative Summary of Significant Associations of Clinical Variables with Taxa and Pathways
This table is too large to be displayed, and is located in an appended Excel file. This qualitatively summarizes statistically significant ($p<0.05$) relationships between clinical variables and microbiome features (pathway and taxa). All two-tailed p-values, coefficients, and confidence intervals are from longitudinal maximum-likelihood GLMMs. Full model information for each microbiome feature is in Table S3.

Supplementary Table 6: Sample Size for Binary Variables
This table is too large to be displayed, and is located in an appended Excel file. There was a total of 402 samples included in this study. The number of samples with associated binary variable values of 1, 0, and missing/unknown values are shown.

Supplementary Table 7: Infant Formula Brands and Ingredients
This table is too large to be displayed, and is located in an appended Excel file. This table shows formula brands reported by parents in this study, and carbohydrate, protein, prebiotic, and probiotic ingredients in each formula according to manufacturers' labels. Dark-shaded squares indicate that an ingredient was present. Fats not displayed, as all formulas included soy oil, coconut oil, palm olein, and high-oleic sunflower and/or safflower oils.

Supplementary Table 8: Taxa identified in Zymobiomics Community Standard Positive Control Samples
This table is located in an appended Excel file, and shows the average (+/- standard deviation) relative abundance of taxa identified in positive control samples (Zymobiomics community standard D6300) across all sequencing runs. All taxa are inferred using MetaPhLan2 as described in the main text. *Nauvomozyma unclassified* and *Eremothecium unclassified* were not identified in any fecal samples, and *Pantoea unclassified* was only found in a relatively small number of fecal samples (N=72 out of 402). There were no taxa found in negative control samples. Positive and negative controls did not suggest any evidence of systemic contamination.