

## **Supplementary Online Content**

Madan JC, Hoen AG, Lundgren SN, et al. Association of cesarean delivery and formula supplementation with the intestinal microbiome of 6-week-old infants. *JAMA Pediatr*. Published online January 11, 2016. doi:10.1001/jamapediatrics.2015.3732.

**eAppendix.** Supplementary Methods

**eReferences**

**eTable.** Subjects by Delivery Mode and Feeding Method

This supplementary material has been provided by the authors to give readers additional information about their work.

## eAppendix. Supplementary Methods

**Targeted 16S rRNA gene sequencing.** 16S rDNA V4-V5 amplicons were generated from purified genomic DNA samples using fusion primers. Forward primers contained one of eight five-nucleotide barcodes between the Illumina-specific bridge and sequencing primer regions and the 16S-specific region. The single reverse primer contained one of 12 Illumina indices. The combination of forward and reverse primers results in 96 unique barcode-index combinations, thus allowing multiplexing of up to 96 samples per lane. PCR was carried out in triplicate 33 µL reaction volumes with an amplification cocktail containing 1.0 U Platinum Taq Hi-Fidelity Polymerase (Life Technologies), 1X Hi-Fidelity buffer, 200 µM dNTP3 PurePeak DNA polymerase mix (Pierce Nucleic Acid Technologies), 1.5 mM MgSO<sub>4</sub> and 0.2 µM of each primer. Each library pool was quantitated by qPCR (Kapa Biosystems) and sequenced on one lane of an Illumina MiSeq 100 cycle paired end run. Datasets were demultiplexed by index using CASAVA 1.8.2 and divided using a custom script. Assignment of pyrotags was used to identify bacterial phylotypes using the GAST pipeline<sup>1-3</sup>, which compares pyrotags to known rRNA genes that have been placed in a phylogenetic framework of more than 1,000,000 nearly full-length rRNA reference sequences (RefSSU) based on the SILVA database<sup>4</sup>. Sequence data are stored in GENBANK: <http://www.ncbi.nlm.nih.gov/genbank/> under accession number PRJNA296814. Internal quality control was performed at MBL by performing PCR reactions in triplicate with one negative control.

**Microbiome profiling.** To generate microbial population structure profiles, we used full-length amplicon sequences for the ribosomal RNA gene V4-V5 hypervariable regions. Our quality control procedures eliminated sequences containing more than one ambiguous nucleotide, removed sequences with a length outside of the expected distribution, and eliminated chimeric reads using the UCHIME algorithm *de novo* and with reference within the USEARCH program<sup>5</sup>. To estimate diversity and perform taxon-independent analyses, we used a pre-clustering strategy<sup>6</sup> that avoids inflated estimates of OTUs.

**Data analysis.** We evaluated the associations of feeding method and delivery mode with microbiome community composition using generalized UniFrac analysis<sup>7</sup>. To avoid sequencing depth bias, the OTU table was rarefied to the minimum sequencing depth in the data set prior to analysis. The phylogenetic tree required for UniFrac analysis was computed using FastTree<sup>8</sup> and was midpoint rooted. Statistical comparisons of groups were made with a permutational multivariate analysis of variance using distance matrices method using the adonis function in the R package “vegan” with 10,000 permutations. Clusters were visualized using principal coordinates analysis of the UniFrac distances. We computed mean pairwise generalized UniFrac distances for pairs both within and between delivery mode groups (vaginal or cesarean section) and feeding method groups (exclusively breastfed, exclusively formula fed or combination fed). Statistical comparisons of mean UniFrac distances were performed using the Kruskal-Wallis rank sum test. We corrected for multiple comparisons in pairwise comparisons of by controlling the false discovery rate. For each hypothesis test we calculated a q-value, the false discovery rate analogue to the p-value<sup>9</sup>. We report both p-values and q-values, and assigned significance to p<0.05 and q<0.1.

Genus-specific read counts were normalized by dividing by the total number of reads for that sample, and normalized read counts for each taxon represented a measure of the relative abundance of the various taxa identified in that sample. We fit linear models to individual log-

ratio transformed<sup>10</sup> relative abundance values to identify the specific genera associated with either feeding mode or delivery method. The log-ratio transformation addresses the constraint that each subject's microbiome profile is composed of a series of observations represented by proportions of a whole, which can introduce dependence between variables that can lead to spurious associations. Zero counts were handled using the zero-replacement procedure described by Aitchison<sup>10</sup>. For each bacterial taxon, we fit a linear model with categorical terms for delivery mode and feeding method. Sample size considerations precluded evaluating interactions between these two variables. Model contrasts were set so that positive delivery mode-associated coefficients can be interpreted as representing a positive association with vaginal delivery, and positive feeding method-associated regression coefficients can be interpreted as representing a positive association with breast milk exposure. Multiple comparisons were addressed by controlling the false discovery rate as described above.

## eReferences

1. Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Relman DA, Sogin ML. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet.* Nov 2008;4(11):e1000255.
2. Huber JA, Mark Welch DB, Morrison HG, et al. Microbial population structures in the deep marine biosphere. *Science.* Oct 5 2007;318(5847):97-100.
3. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* 2007;8(7):R143.
4. Pruesse E, Quast C, Knittel K, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 2007;35(21):7188-7196.
5. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimeric detection. *Bioinformatics.* Aug 15 2011;27(16):2194-2200.
6. Huse SM, Mark Welch D, Morrison HG, Sogin ML. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol.* Jul 2010;12(7):1889-1898.
7. Chen J, Bittinger K, Charlson ES, et al. Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics.* Aug 15 2012;28(16):2106-2113.
8. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol.* Jul 2009;26(7):1641-1650.
9. Storey JD. The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann Stat.* Dec 2003;31(6):2013-2035.
10. Aitchison J. *The statistical analysis of compositional data.* London ; New York: Chapman and Hall; 1986.

**eTable.** Subjects by Delivery Mode and Feeding Method (N=102)

	<b>Vaginal</b>	<b>Cesarean section</b>
<b>Exclusively breastfed</b>	50	20
<b>Combination</b>	16	10
<b>Exclusively formula-fed</b>	4	2
Fisher exact test of association: p=0.66		