## Supplemental Material and Methods

**DNA extraction for 16S rRNA gene amplification and sequencing:**

DNA isolation by mechanic and enzymatic lysis was performed by MagNA Pure LC DNA Isolation Kit III (Bacteria, Fungi) (Roche, Mannheim, Germany) in MagNA Lyser green beads tubes (Roche, Mannheim, Germany) in a MagNA Lyser Instrument (Roche, Mannheim, Germany) according to Klymiuk et al.[1]. For PCR, Primers targeting hypervariable region V4 were used for amplification according to the 16S Illumina Amplicon Protocol of the Earth Microbiome Project[2] (515F[3] FWD:GTGYCAGCMGCCGCGGTAA; 806R[4] REV:GGACTACNVGGGTWTCTAAT), were used.

**DNA extraction for metagenomic analysis and shotgun sequencing:**

In a subset of available stool samples we performed shotgun sequencing. We matched sFPIP infants with patients form the control group for age, mode of delivery, sex and nutrition at baseline. We performed shotgun sequencing for 60 fecal samples in total. Reduced sample numbers in the sFPIP group were caused by sample loss due to prior 16S analysis or extraction failures: control: week 0 n=12 / week 4 n=12 / week 8 n=12; sFPIP: week 0 n=11 / week 4 n=7 / week 8 n=6).

DNA isolation (Shotgun analysis) Extraction, lysis and DNA isolation was done according to manufacturer's recommendation (Fast DNA Stool Mini Kit Qiagen). Bead beating was run on a fastprep24 instrument (MPBiomedicals; 4 cycles of 45s at speed 4) in 2ml screwcap tubes containing 0.6g 0.1mm glass beads. 200µl of raw extract was prepared for DNA-isolation. Concentration of the isolated DNA was assessed with PicoGreen measurement (Quant-iT™ PicoGreen™ dsDNA Assay Kit, Thermo Fisher) and integrity was checked by agarose gel electrophoresis for a random sample.

Illumina library preparation and sequencing Illumina TruSeq DNA libraries were prepared using the TruSeq nano DNA Library preparation kit (Illumina, San Diego, USA) according to the manufacturer’s instructions. Subsequently the libraries were checked for quality and

library size on a 2100 Bioanalyzer instrument using a High Sensitivity DNA Assay kit (Agilent Technologies, Santa Clara, USA). The final libraries were quantified using a Quant-iT™ PicoGreen™ ds DNA Assay Kit (Thermo Fisher Scientific, Waltham, USA) and equimolarly pooled prior to sequencing. Sequencing was performed on an Illumina MiSeq sequencing system using a MiSeq Reagent Kit v2 for 500 cycles (Illumina, San Diego, USA). Total past filter bases: 338,683,403,141 Mean read length: 150 bp Q20 %: 96 Q30 %: 92 Mean Q: 34 Masked: 596,991

**Computational Methods for Microbiome Analysis**

For analysis of 16S amplicon metagenomics sequences were imported to Qiime2 (v. 2019.4)[5]. Quality control was performed on 14,518,615 demultiplexed sequences total (control: 62,326±1,325; sFPIP: 67,672±1,784 (mean±SEM)) using DADA2 (denoise paired)[6], to improve the overall sequence quality. First 10 nucleotides were trimmed and reads were truncated at nucleotide 282 (forward read) or 206 (reverse read) respectively. To assign taxonomy to our sequences we used a pre-trained Naïve Bayes that was trained on Silva 132 99% OTUs[7], where sequences have been trimmed to only include sequences from the V4 region that was used for 16S amplicon sequencing according to the Earth Microbiome Project[2]. Mitochondria and Chloroplast reads were excluded from feature tables. Samples below 10,000 reads were removed from analysis. Alpha (Shannon, evenness, Observed otus, Faith PD) and Beta diversity (Bray Curtis, Jaccard, weighted and unweighted UniFrac) were calculated to compare microbial communities using Qiime's plugin qiime diversity core-metrics-phylogenetic. For further differential abundance analysis we implemented LefSe analysis[8] (Galaxy Version 1.0), qiime2 plugins for ALDEx2[9], ANCOM[10], DEICODE[11] and Qurro[12]. Further data analysis was performed in R[13] using qiime2R package[14].

For taxonomic profiling quality control was performed on shotgun sequences using KneadData (0.7.2)[15] and subsequently analyzed with MetaPhlAn2 to profile the structure and the composition of microbial communities at species resolution.

The datasets generated and analysed during the current study are available in the European nucleotide archive (ENA) repository under the Primary accession number PRJEB45874 (Secondary accession ERP130046).


1. Klymiuk I, Bilgilier C, Stadlmann A, Thannesberger J, Kastner MT, Hogenauer C, et al. The Human Gastric Microbiome Is Predicated upon Infection with Helicobacter pylori. *Front Microbiol*. 2017;8:2508.
2. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*. 2017;551(7681):457-463.
3. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol*. 2016;18(5):1403-1414.
4. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology*. 2015;75(2):129-137.
5. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37(8):852-857.
6. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature methods*. 2016;13(7):581-583.
7. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41(Database issue):D590-596.
8. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12(6):R60.
9. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome*. 2014;2:15.
10. Mandal S, Van Treuren W, White RA, Eggesbo M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb Ecol Health Dis*. 2015;26:27663.
11. Martino C, Morton JT, Marotz CA, Thompson LR, Tripathi A, Knight R, et al. A Novel Sparse Compositional Technique Reveals Microbial Perturbations. *mSystems*. 2019;4(1).
12. Fedarko MW, Martino C, Morton JT, Gonzalez A, Rahman G, Marotz CA, et al. Visualizing 'omic feature rankings and log-ratios using Qurro. *NAR Genom Bioinform*. 2020;2(2):lqaa023.
13. Team RC. R: A language and environment for statistical computing. http://www.R-project.org/. Published 2013.
14. Bisanz JE. qiime2R: Importing QIIME2 artifacts and associated data into R sessions. https://github.com/jbisanz/qiime2R. Published 2018.
15. https://github.com/biobakery/kneaddata.