Methods

Cohort description and data collection

The study subjects are children from the FinnBrain Cohort Study (36) that is a general population birth cohort study located in the southwestern Finland. The FinnBrain Birth Cohort Study recruited families with sufficient fluency in Finnish or Swedish, and normal 1st trimester ultrasound examination. A subset of the cohort participated in the study visits, and there were no exclusion criteria for the collection of fecal samples. The initial recruitment took place between December 2011 and April 2015, and fecal samples were collected from May 2013 to May 2018. The fecal samples were collected from the children by the parents according to written and oral instructions at 2.5, 6, 14 and 30 months postpartum. The samples were collected in plastic tubes, and parents were instructed to store the sample in a refrigerator, and bring the sample to the laboratory within 24 h. The sample collection time was reported. Clinical data used in the study were collected with parental reports during and after pregnancy at 14, 24, 34 gestational weeks, 3, 6, 12, and 24 months postpartum and during study visits (2.5, 6, 14, and 30 months). Likewise, the data on maternal pre-pregnancy body mass index (BMI; kg/m2), duration of gestation as well as mode of delivery (caesarian section vs. vaginal) were collected from National Birth Registry provided by the National Institute for Health and Welfare of Finland (www.thl.fi). The information on maternal perinatal and infant neonatal intravenous antibiotic intake was collected from the hospital records. Breastfeeding was categorized in two ways: 1) any current breastfeeding (yes vs. no); 2) exclusive breastfeeding at least 4 months and partial breastfeeding for at least 6 months (breastfeeding_criteria, yes vs. no). The Ethical Committee of Southwestern Finland approved the study. Parents provided informed consent on behalf of their children. STORMS guideline was used for reporting the methods and materials.

Metabolome analysis

The BAs were measured in fecal samples as described previously (16). Only samples frozen within 24 h of sample collection were included in the metabolome analyses. The order of the samples was randomized before sample preparation. Two aliquots (50 mg) of each fecal sample were weighed. An aliquot was freeze-dried prior to extraction to determine the dry weight. The second aliquot was homogenized by adding homogenizer beads and 20 µL of water for each mg of dry weight in the fecal sample, followed by samples freezing to at least -70 °C and homogenizing them for five minutes using a bead beater. The BAs analysed were Litocholic acid (LCA), 12-oxo-litocholic acid(12-oxo-LCA), Chenodeoxycholic acid (CDCA), Deoxycholic acid (DCA), Hyodeoxycholic acid (HDCA), Ursodeoxycholic acid (UDCA), Dihydroxycholestanoic acid (DHCA), 7-oxo-deoxycholic acid (7-oxo-DCA), 7-oxohyocholic acid (7-oxo-HCA), Hyocholic acid(HCA), β-Muricholic acid (b-MCA), Cholic acid (CA), Ω/α-Muricholic acid (w/a-MCA), Glycolitocholic acid (GLCA), Glycochenodeoxycholic acid (GCDCA), Glycodeoxycholic acid (GDCA), Glycohyodeoxycholic acid (GHDCA), Glycoursodeoxycholic acid (GUDCA), Glycodehydrocholic acid (GDHCA), Glycocholic acid (GCA), Glycohyocholic acid (GHCA), Taurolitocholic acid (TLCA), Taurochenodeoxycholic acid (TCDCA), Taurodeoxycholic acid (TDCA), Taurohyodeoxycholic acid (THDCA), Tauroursodeoxycholic acid (TUDCA), Taurodehydrocholic acid (TDHCA), Tauro-αmuricholic acid (TaMCA), Tauro-β-muricholic acid (TbMCA), Taurocholic acid (TCA), Trihydroxycholestanoic acid (THCA) and Tauro-Ω-muricholic acid (TwMCA). BAs were extracted by adding 40 µL fecal homogenate to 400 µL crash solvent (methanol containing 62,5 ppb each of the internal standards LCA-d4, TCA-d4, GUDCA-d4, GCA-d4, CA-d4,

UDCA-d4, GCDCA-d4, CDCA-d4, DCA-d4 and GLCA-d4) and filtering them using a Supelco protein precipitation filter plate. The samples were dried under a gentle flow of nitrogen and resuspended using 20 µL resuspenstion solution (Methanol:water (40:60) with 5 ppb Perfluoro-n-[13C9]nonanoic acid as in injection standard). Quality control (QC) samples were prepared by combining an aliquot of every sample into a tube, vortexing it and preparing QC samples in the same way as the other samples. Blank samples were prepared by pipetting 400 µL crash solvent into a 96-well plate, then drying and resuspending them the same way as the other samples. Calibration curves were prepared by pipetting 40 µL of standard dilution into vials, adding 400 µL crash solution and drying and resuspending them in the same way as the other samples. The concentrations of the standard dilutions were between 0.0025 and 600 ppb. The LC separation was performed on a Sciex Exion AD 30 (AB Sciex Inc., Framingham, MA) LC system consisting of a binary pump, an autosampler set to 15 °C and a column oven set to 35 °C. A waters Aquity UPLC HSS T3 (1.8µm, 2.1x100mm) column with a precolumn with the same material was used. Eluent A was 0.1 % formic acid in water and eluent B was 0.1 % formic acid in methanol. The gradient started from 15 % B and increased to 30 % B over 1 minute. The gradient further increased to 70 % B over 15 minutes. The gradient was further increased to 100 % over 2 minutes. The gradient was held at 100 % B for 4 minutes then decreased to 15 % B over 0.1 minutes and re-equilibrated for 7.5 minutes. The flow rate was 0.5 mL/min and the injection volume was 5 µL. The mass spectrometer used for this method was a Sciex 5500 QTrap mass spectrometer operating in scheduled multiple reaction monitoring mode in negative mode. The ion source gas1 and 2 were both 40 psi. The curtain gas was 25 psi, the CAD gas was 12 and the temperature was 650 °C. The spray voltage was 4500 V. Data processing was performed on Sciex MultiQuant.

Quantification of SCFA

We adapted and modified the targeted SCFA analysis from previous work (37). Fecal samples were homogenized by adding water (10 µL per mg of dry weight as determined for the BA analysis) to wet feces, the samples were homogenized using a bead beater. Analysis of SCFA was performed on fecal homogenate (50 µL) crashed with 500 µL methanol containing internal standard (propionic acid-d6 and hexanoic acid-d3 at 10 ppm). Samples were vortexed for 1 min, followed by filtration using 96-Well protein precipitation filter plate (Sigma-Aldrich, 55263-U). Retention index (RI, 8 ppm C10-C30 alkanes and 4 ppm 4,4Dibromooctafluorobiphenyl in hexane) was added to the samples. Gas chromatography (GC) separation was performed on an agilent 5890B GC system equipped with a Phenomenex Zebron ZB-WAXplus (30 m × 250 µm × 0.25 µm) column a short blank precolumn (2 m) of the same dimensions was also added. A sample volume of 1 µL was injected into a split/splitless inlet at 285°C using split mode at 2:1 split ratio using a PAL LSI 85 sampler. Septum purge flow and split flow were set to 13 mL/min and 3.2 mL/min, respectively. Helium was used as carrier gas, at a constant flow rate of 1.6 mL/min. The GC oven program was as follows: initial temperature 50°C, equilibration time 1 min, heat up to 150°C at the rate of 10°C/min, then heat at the rate of 40°C/min until 230°C and hold for 2 min. Mass spectrometry was performed on an Agilent 5977A MSD. Mass spectra were recorded in Selected Ion Monitoring (SIM) mode. The detector was switched off during the 1 min of solvent delay time. The transfer line, ion source and quadrupole temperatures were set to 230, 230 and 150°C, respectively. Dilution series of SCFA standards of acetic, propionic, butyric, valeric, hexanoic acid, isobutyric, and iso-valeric acid were prepared in concentrations of 0.1, 0.5, 1, 2, 5, 10, 20, 40, and 100 ppm for the construction of standard curves for quantification.

Analysis of polar metabolites

Polar metabolites were extracted in methanol. The method was adapted from the method used by Lamichhane et al. (24). Fecal homogenate (60 µL) were diluted with 600 µL methanol crash solvent containing internal standards (heptadecanoic acid (5 ppm) valine-d8 (1 ppm) and glutamic acid-d5 (1 ppm)). After precipitation the samples were filtered using Supelco protein precipitation filter plates. One aliquot (50 µL) was transferred to a shallow 96-well plate to create a QC sample. The rest of the sample volume was dried under a gentle stream of nitrogen and stored in -80 °C until analysis. After thawing the samples were again dried to remove any traces of water. Derivatization was carried out on a Gerstel MPS MultiPurpoe Sampler using the following protocol: 25 µL methoxamine (20 mg/mL) was added to the sample followed by incubation on a shaker heated to 45 °C for 60 minutes. N-Methyl-N(trimethylsilyl) trifluoroacetamide (25 µL) was added followed by incubation (60 min). After that, 25 µL retention index was added, the sample was allowed to mix for one min followed by injection. The automatic derivatization was carried out using the Gerstel maestro 1 software (version 1.4). Gas chromatographic (GC) separation was carried out on an Agilent 7890B GC system equipped with an Agilent DB-5MS (20 m x 0,18 mm (0,18 µm)) column. A sample volume of 1 µl was injected into a split/splitless inlet at 250°C using splitless mode. The system was guarded by a retention gap column of deactivated silica (internal dimensions 1.7 m, 0.18 mm, PreColumn FS, Ultimate Plus Deact; Agilent Technologies, CA, USA). Helium was used as carrier gas at a flow rate of 1.2 ml/min for 16 min followed by 2 mL/min for 5.75 min. The temperature programme started at 50°C (5 min), then a gradient of 20°C/min up to 270°C was applied and then finally a gradient of 40°/min to 300°C, where it was held stable for 7 min. The mass spectrometry was carried out on a LECO Pegasus BT system (LECO). The acquisition delay was 420 sec. The acquisition rate was 16 spectra/sec. The mass range was 50 – 500 m/z and the extraction frequency was 30 kHz. The ion source was held at 250 °C and the transferline heater temperature was 230 °C. ChromaTOF software (version 5.51) was used for data aquisition. The samples were run in 9 batches, each consisting of 100 samples and a calibration curve. In order to monitor the run a blank, a QC and a standard sample with a known concentration run between every 10 samples. Between every batch the septum and liner on the GC were replaced, the precolumn was cut if necessary and the instrument was tuned. The retention index was determined with ChromaTOF using the reference method function. For every batch a reference file was created. The reference file contained the spectras and approximate retention times of the alkanes from C10 to C30 as determined manually). A reference method was implemented for every sample in order to determine the exact retention time of the alkanes. Text files with the names and retention times of the alkanes were then exported and converted to the correct format for MSDIAL using an in-house R script. The samples were exported from ChromaTOF using the netCDF format. After this they were converted to abf files using the abfConverter software (Reifycs). Untargeted data processing was carried out using MSDIAL (version 4.7). The minimum peak height was set to an amplitude of 1000, the sigma window value was 0.7 and the EI spectra cut off was 10. The identification was carried out using retention index with the help of the GCMS DBPublic-kovatsRI-VS3 library provided on the MSDIAL webpage. A separate RI file was used for each sample. The RI tolerance was 20 and the m/z tolerance was 0.5 Da. the EI similarly cut off was 70 %. The identification score cut off was 70 % and retention information was used for scoring. Alignment was carried out using the RI with an RI tolerance of 10. The EI similarity tolerance was 60 %. The RI factor was 0.7 and the EI similarity factor was 0.5. The results were exported as peak areas and further processed with excel. In excel the results were normalized using heptadecanoic acid as internal standard and the features with a coefficient of variance of less than 30 % in QC samples

were selected. Further filtering was carried out to remove alkanes and duplicate features. The IDs of the features which passed the CV check were further checked using the Golm Metabolome Database.

Microbiome analysis

DNA extraction and sample processing

The samples were divided into cryotubes and freezed in -80C within 2 days after arriving at the laboratory. Samples were kept at +4C before freezing. Only samples that were freezed within 48 h of sample collection were sequenced. Sample volume for DNA extraction was approximately 100 mg. Lysis buffer was added 1 ml and the samples were homogenized with glass beads 1000 rpm / 3 min. The samples were centrifuged at high speed (> 13000 rpm) for 5 min. The lysate (800µL) was then transferred to tubes and the extraction proceeded according to the manufacturer's protocol. DNA was extracted using a semiautomatic extraction instrument Genoxtract with DNA stool kit (HAIN life science, Germany).

DNA yields were measured with Qubit fluorometer using Qubit dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific, USA). The DNA extraction and sequencing was performed in the University of Turku. 16S ribosomal RNA (rRNA) amplicon sequencing Bacterial community composition was determined by sequencing the V4 region of 16S rRNA gene using Illumina MiSeq platform (Illumina, USA). The sequence library was constructed with an in-house developed protocol where amplicon PCR and index PCR were combined. The DNA samples were diluted in PCR grade water to 10 ng/µL concentration prior to library PCR. PCR was performed with KAPA HiFi High Fidelity PCR kit with dNTPs (Roche, USA). Reverse and forward primers included in-house modifications verified by Rintala et al. (38). The forward and reverse primer sequences were 5'-AATGATACGGCGACCACCGAGATCTACAC -i5- TATGGTAATT -GTGTGCCAGCMGCCGCGGTAA-3' and 5'-CAAGCAGAAGACGGCATACGAGAT -i7AGTCAGTCAG-GC-GGACTACHVGGGTWTCTAAT-3', respectively, where i5 and i7 indicate the sample specific indexes. After PCR, 5µl of the product was analyzed with 1,5% TBE agarose gel (100V, 1h15min). PCR products were purified with AMPure XP magnetic beads (Becman Coulter, USA). The DNA concentration of the purified samples were measured with Qubit fluorometer using Qubit dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific, USA), after which the samples were mixed in equimolar concentration into a 4 nM library pool. The library pool was denatured, diluted to a concentration of 4 pM and a denaturized PhiX control (Illumina, USA) was added. The sequencing was performed with Illumina MiSeq Reagent kit v3 (600 cycles) on MiSeq system with 2x 250 base pair (bp) paired ends following the manufacturer's instructions. Positive control (DNA 7-mock standard) and negative control (PCR grade water) were included in library preparation and sequencing runs (Supplementary figures 5-8). DADA2-pipeline (version 1.14) was used to preprocess the 16S rRNA gene sequencing data to infer exact amplicon sequence variants (ASVs) (39). The reads were truncated to length 225 and reads with more than two expected errors were discarded (maxEE = 2). SILVA taxonomy database (version 138) (40,41) and RDP Naive Bayesian Classifier algorithm (42) were used for the taxonomic assignments of the ASVs.

Statistical analyses

The data analyses were performed with R version 4.2.0 with packages including phyloseq, mia, vegan, DirichletMultinomial and lme. Heatmaps were created with the pheatmap R package. Shannon Index and Inverse Simpson were used as alpha diversity indices and those were calculated

with mia package from the untransformed ASV-table. Metabolite concentrations were log-transformed with a pseudocount (minimum value / 2). Dirichlet Multinomial Mixtures model with the rarefied, genus-level count data were used for clustering the microbiome data. The number of community types was justified by the Laplace criteria. Factor analysis, the relative contribution of a clinical/demographic factor towards the total variance of the metabolite classes were estimated by fitting a linear regression model. The total metabolite concentrations of a particular class was regressed to a clinical/demographic factor of interest, and median marginal coefficient of determination (R2) and % of explained variance were estimated. Factor analysis was performed using the 'Scater' package deployed in R. Wilcoxon test, Chi-square test, and Kruskal-Wallis test with Dunn's posthoc test were used in the analyses. Linear mixed models with child ID as random effect and sampling age as fixed effect was used to study i. metabolite age-trends, ii. association between metabolite concentrations and demographic factors, iii., association between microbiome community typemembership and demographic factors, iv. associations between metabolite concentrations and microbiome community type membership, and v. association between metabolite concentrations and the interaction with breastfeeding and rclr-transformed prevalent genera abundances as breastfeeding has been shown to drive the microbiome maturation (4). Package lme4 was used to check for model singularity, and nlme was used for running the mixed model. The clr-module from the ALDEx2 package was used for the differential abundance analysis (43). Variance explained in the metabolome assays by demographic factors was calculated with the package scater (44). p-Values were adjusted for multiple testing with Benjamini-Hochberg procedure.

References

4 Stewart, C. J. et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. Nature 562, 583-588, doi:10.1038/s41586-018-0617-x (2018).

16 Lamichhane, S., Sen, P., Dickens, A. M., Oresic, M. & Bertram, H. C. Gut metabolome meets microbiome: A methodological perspective to understand the relationship between host and microbe. Methods 149, 3-12, doi:10.1016/j.ymeth.2018.04.029 (2018).

24 Lamichhane, S. et al. Dysregulation of secondary bile acid metabolism precedes islet autoimmunity and type 1 diabetes. Cell Rep Med 3, 100762, doi:10.1016/j.xcrm.2022.100762 (2022).

36 Karlsson, L. et al. Cohort Profile: The FinnBrain Birth Cohort Study (FinnBrain). Int J Epidemiol 47, 15-16j, doi:10.1093/ije/dyx173 (2018).

37 Trimigno, A. et al. Identification of weak and gender specific effects in a short 3 weeks intervention study using barley and oat mixed linkage beta-glucan dietary supplements: a human fecal metabolome study by GC-MS. Metabolomics 13, 108, doi:10.1007/s11306-017-1247-2 (2017).

38 Rintala, A. et al. Gut Microbiota Analysis Results Are Highly Dependent on the 16S rRNA Gene Target Region, Whereas the Impact of DNA Extraction Is Minor. J Biomol Tech 28, 19-30, doi:10.7171/jbt.17-2801-003 (2017).

39 Callahan, B. J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13, 581-583, doi:10.1038/nmeth.3869 (2016).

40 Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41, D590-596, doi:10.1093/nar/gks1219 (2013).

41 Yilmaz, P. et al. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res 42, D643-648, doi:10.1093/nar/gkt1209 (2014).

42 Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73, 5261-5267, doi:10.1128/AEM.00062-07 (2007).

43 Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G. & Gloor, G. B. ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. PLoS One 8, e67019, doi:10.1371/journal.pone.0067019 (2013).

44 McCarthy, D. J., Campbell, K. R., Lun, A. T. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics 33, 1179-1186, doi:10.1093/bioinformatics/btw777 (2017).