SUPPLEMENTAL INFORMATION

EXTENDED EXPERIMENTAL PROCEDURES

Human studies

Human studies were conducted with approval of institutional review boards from the University of Malawi, Pirkanmaa Hospital (Finland) and Washington University School of Medicine in St. Louis. The mothers and infants included in this analysis were enrolled in two randomized, controlled, single-blind, parallel group clinical trials, LCNI-5 and iLiNS-DYAD-M [ClinicalTrials.gov identifiers: NCT00524446 and NCT01239693] of micronutrient fortified lipid-based nutrient supplements conducted in southern Malawi (Ashorn et al., 2015; Mangani et al., 2013). Anthropometric measurements of 6 month-old infants were performed in triplicate to determine height-for-age Z-scores (HAZ).

De-identified breast milk samples were collected from women enrolled in the two trials when their infants were 6 months old. The entire contents of one breast were manually expressed into a sterile container approximately 2 h after the last reported feeding on that breast. Aliquots were stored at -40°C to -80°C in Malawi and later shipped on dry ice to Washington University in St. Louis, where the samples were maintained at −80 °C until being shipped to UC Davis for HMO analysis. A fresh morning fecal sample was collected from the infants of these mothers on the day of their 6-month clinic visit. The sample was divided into aliquots in a laboratory and stored at −20°C on the same day. Within three days, samples were transferred to -40°C or -80°C freezers in Malawi prior to shipment to Washington University.

Mass spectrometry of breast milk samples

Mass spectrometry analysis of breast milk HMO content was performed on all available samples from mothers of children enrolled in the LCNI-5 and iLiNS-DYAD-M study cohorts that met the anthropometric criteria stated in the main text (n=88 and 215 for LCNI-5 and iLiNS,

respectively). Free HMOs from whole milk were extracted using previously reported methods (Ninonuevo et al., 2006; Wu et al., 2010; Wu et al., 2011; Totten et al. 2014). Briefly, 50 μL aliquots of breast milk samples obtained from mothers in the LCNI-5 cohort and 25 μL aliquots of samples obtained from members of the iLiNS-DYAD-M cohort were distributed into 96-well plates, diluted and defatted by centrifugation. To precipitate proteins, two volumes of ethanol were added to the aqueous layer and the mixtures incubated for 1.5 hours at -80°C. After centrifugation, the supernatant was collected, and the resulting glycans were reduced with 1.0 M $NaBH_4$ for 1.5 hours at 65°C. Samples were then purified by solid phase extraction, loaded onto graphitized carbon cartridges, desalted with nanopure water and then eluted with 20% acetonitrile in water and 40% acetonitrile in 0.05% trifluoroacetic acid (v/v). The eluent fractions were combined and the solvent evaporated. After reconstitution, the samples were diluted to appropriate concentrations and subjected to nano-high performance liquid chromatography (HPLC)-chip/time-of-flight (TOF) mass spectrometry.

The Agilent 1200 series liquid chromatography unit utilizes a dual pump system with sample loading and analyte separation performed on a microfluidic chip. The chip has a 40 nL enrichment column and a 75 μL x 43 mm analytical column packed with porous graphitized carbon. This system is coupled to an Agilent 6220 series TOF mass spectrometer via chip-cube interface. The HPLC capillary pump loads the sample onto the chip's enrichment column at a flow rate of 4.0 μL/min with a 1 μL injection volume, and the nano pump achieves separation with a binary gradient of aqueous solvent (3% acetonitrile/water (v/v) in 0.1% formic acid (FA) and organic solvent (90% acetonitrile/water (v/v) in 0.1% FA). The separation method was developed and optimized for HMO mixtures (Wu et al., 2010; Wu et al., 2011). To monitor instrument performance and minimize batch effects, an external HMO pool was run intermittently throughout sample analysis.

Data were collected using Agilent MassHunter Workstation Data Acquisition software, version B.02.01 and then analyzed using Agilent MassHunter Qualitative Analysis software version B.03.01. HMO compounds were identified using the 'Find Compounds by Molecular Feature' function and peak alignment was performed using custom scripts. Specific structures were assigned by matching exact mass (within 20 ppm mass error of theoretical values) and retention time to established HMO libraries (Wu et al., 2010; Wu et al., 2011). Absolute abundances in ion counts were normalized to total HMO abundance found in the external HMO pool run respective to that sample's batch. Glycan types were divided into four classes: fucosylated, sialylated, fucosylated and sialylated, and non-fucosylated neutrals. Relative class abundances were calculated by dividing absolute class abundance by each mother's total HMO abundance. The same normalization and relative abundance calculations were performed for individual structures.

Phenotypic secretor status for all mothers was determined by analyzing abundances of individual HMO structures and examining their chromatographic profiles. LCNI-5 secretor status assignment was based on a previously developed method (Totten et al., 2012). Secretor status assignment for mothers in the iLiNS-DYAD-M cohort also used this method but with the following modifications: absolute abundances of compounds with known α1-2 fucose linkage [fucose(α1-2)galactose] were summed and normalized to each mother's total oligosaccharide abundance in order to calculate relative α1-2 fucosylation for each mother. The structures used for this determination were 2'fucosyllactose (2'FL), lactodifucotetraose (LDFT), trifucosyllacto-*N*-hexaose (TFLNH), difucosyllacto-*N*-hexaose a (DFLNHa) and fucosyl-*iso*-lacto-*N*-hexaose I (IFLNH I) (Newburg and Neubauer, 1995). Secretor status was confirmed by comparing the extracted ion chromatograms of those structures between assigned secretors and non-secretors.

Generating a clonally arrayed sequenced collection of bacterial strains from a 6-month old severely stunted Malawian infant

A de-identified fecal sample, collected and immediately frozen in a dry cryogenic storage dewar (pre-charged with liquid nitrogen), from a 6-month old infant that had been enrolled in a previously reported study of twins living in five rural southern Malawian villages (patient ID h264A in Yatsunenko et al., 2012), was used to produce a clonally arrayed collection of anaerobic bacterial strains with a protocol modified slightly from what we had described previously (Faith et al., 2013; Goodman et al., 2011; Ridaura et al., 2013). An aliquot of the frozen sample was pulverized in a Biosafety Class II hood with a ceramic mortar and pestle filled with liquid nitrogen. An aliquot (1g) of the pulverized material, sealed in a sterile screw-capped tube (Axygen SCT-200-C-S), was brought into an anaerobic Coy chamber (atmosphere: 20% CO_2, 5% H_2, and 75% N_2), immediately suspended in 15 mL Gut Microbiota Medium (GMM; Goodman et al., 2011) and blended with a Waring MC1 blender set on maximum speed (four cycles of blending for 20 seconds followed by a 30 second pause). The sample was allowed to stand for 5 minutes so that particulate matter could settle by gravity; the resulting supernatant was passed through a 100 μm pore diameter filter (BD systems, Inc. Franklin Lakes, NJ) to remove remaining particulate material, mixed with an equal volume of pre-reduced GMM containing 30% glycerol (final concentration 15% glycerol) and placed in Wheaton crimp top tubes for storage at -80°C.

A vial of the -80°C anaerobic glycerol stock containing an aliquot of the clarified sample was diluted into pre-reduced GMM lacking resazurin in an anaerobic chamber and dispensed into a pre-reduced 384-well flat-bottomed polypropylene multi-well plate (170 μL per well; Costar #3964, Corning, NY). All liquid handling steps were performed using a set of custom

interfaces for a Precision XS robot (BioTek) so that picking, arraying and archiving of the bacterial culture collection could be done with economy and speed under anaerobic conditions.

To determine the dilution at which approximately 30% of wells would receive one or more bacterial cells in the initial inoculation, two- and fourfold serial dilutions of the clarified fecal suspension were performed (from 10^{-6} to 10^{-10}) in a trial inoculation (48 wells per dilution; 170 µL per well). Plates (Nunc 260251) were sealed with sterile foil lids (VWR 60941-076) and incubated in a Coy chamber for 3 days at 37°C under anaerobic conditions (atmosphere: 20% CO_2, 5% H_2, and 75% N_2). The dilution at which ~30% of wells were turbid (OD_{630} > 0.1) was chosen for a subsequent large-scale culturing effort. To do this, a second vial of the frozen clarified fecal sample was added to 500 mL of pre-reduced GMM at the calculated dilution and dispensed into 384-well plates as above (170 µL per well, 3,840 wells total) and incubated at 37°C under anaerobic conditions for 3 days. Frozen anaerobic glycerol stocks of each plate were prepared by resuspending cells (by pipetting) and transferring 25 µL aliquots to a pre-reduced 384-well plate containing an equal volume of pre-reduced GMM with 30% glycerol. Glycerol stock plates were frozen on dry ice inside the anaerobic chamber and stored in a -80°C freezer. The anaerobic status of frozen stocks was verified by including resazurin in the storage medium. An additional 60 µL of the resuspended cell culture was taken from each well for measuring OD_{630} and sequencing of bacterial genomic DNA.

Based on the results of sequencing amplicons generated from the V4 region of each isolate's 16S rRNA gene, the library was subsequently condensed to include up to seven representatives of each unique V4 16S rRNA sequence. (Condensing at this step consisted of recovering the frozen 384-well stock plates in GMM, purifying each isolate on GMM agar plates and inoculating fresh 96-well plates with the isolates).

Genomic DNA from each isolate was subjected to shotgun sequencing in multiplex using a HiSeq2000 instrument [101 nt paired-end reads, 44.65±29.32 fold coverage (mean±SD), see **Table S2A**]. Genomes were assembled using Velvet and Velvet Optimizer, version 2.1.7. Genes encoding proteins, tRNAs and rRNAs were annotated with Glimmer3.0, tRNAScan 1.23 and RNAmmer 1.2, respectively.

Shared genome content at the nucleotide level was computed using Nucmer as described in an earlier report (Faith et al., 2013). Shared genome content was calculated between two genomes (A and B) by the following formula,

$$Shared\ Genome\ Content = \frac{X + Y}{len(A) + len(B)} \times 100\%$$

where X is the length of A that aligns to B, Y is the length of B that aligns to A, len(A) is the length of A, and len(B) is the length of B. Isolates were clustered at the level of >96% shared genome content using Nucmer (Delcher, 2002). A single representative of each cluster was selected based on draft genome sequence quality and used to generate the 25-member strain collection.

Each of the selected isolates listed in **Table S2A** was re-purified on LYHBHI agar plates (Sokol et al., 2008) and a full-length 16S rRNA amplicon was generated [50 µL PCR reactions contained 44 µL ReddyMix PCR master mix (Thermo Scientific#AB-0575/LD), 2 µL Forward primer (8F; 10 µM) and 2 µL Reverse primer (1392R; 10 µM); cycling conditions were 95°C for 10 minutes, followed by 35 cycles of 95°C for 45 seconds, 56°C for 45 seconds, and 72°C for 2 minutes, followed by 72°C for 10 minutes]. Amplicons were sequenced using the dideoxy chain termination method (Retrogen, Inc; San Diego, CA). Our designation of a strain as *B. longum* subsp. *infantis* was based on (i) the presence of genes encoding urease subunits (LoCascio et al., 2010) and (ii) the presence of HMO utilization gene clusters containing all four glycoside

hydrolases active on HMO linkages [β-galactosidase, sialidase, fucosidase and hexosaminidase) (Sela et al., 2008); see **Table S2C]**.

The 25-strain library was stored in a "sparse" 96-well format where members were deliberately spaced across the plate in between un-inoculated wells employed as a 'buffer' to monitor potential contamination events. To prepare the strain collection for gavage into germ-free animals, a frozen archived 'stock plate' was thawed at room temperature inside the anaerobic chamber and a 30 μL aliquot of the contents of each well was transferred using a Precision XS robot (BioTek) into 600 μL of either fresh pre-reduced GMM or pre-reduced LYHBHI medium (Sokol et al., 2008) located in the wells of a new 96-well microtiter plate (Nunc, #260251). Following an overnight incubation at 37°C in an anaerobic chamber, cultures were independently assayed for growth (OD_{630}), an equal volume (200 μL) was taken from each well and the aliquots were pooled. An equal volume of pre-reduced PBS containing 30% glycerol was mixed with the pooled cultures, and the mixture was dispensed, in 1 mL aliquots, into Wheaton crimp cap tubes and stored at -80°C until use. Stocks of the pooled material were kept at -80°C and then thawed immediately prior to gavage of gnotobiotic animals.

Purification and compositional analysis of S-BMO

S-BMO was purified from commercial whey permeate after lactose was partially removed by concentration and crystallization. BMO in delactose permeate was first concentrated with a single stage ultrafiltration membranes (molecular mass cut off of 1000 Da; GE Waters, Inc.; Boulder, CO, USA) at 30°C. S-BMO was then extracted from this ultrafiltrate by anion-exchange chromatography on a column containing 100 L of dimethylamine functionalized chloromethylated copolymer of styrene and divinyl benzene (Bucher-Alimentech Ltd.; Auckland, NZ). Chromatography was run at 150 liter/hour at 30°C. The S-BMO peak collected from anion-exchange chromatography was further concentrated by

ultrafiltration at 30°C using the same type of membrane to remove minerals in the solution. The final concentrate was freeze-dried and stored in a vacuum desiccator at room temperature.

The total carbohydrate composition of S-BMO was determined on an Agilent 6520 accurate-mass Q-TOF LC/MS with a microfluidic nano-electrospray chip and high-performance anion-exchange chromatography with pulsed amperometric detection (Thermo Scientific HPAE-PAD ICS-5000, Sunnyvale, CA, USA).

Design and preparation of the Malawi 8 (M8) diet

This diet was formulated based on a dietary survey of the complementary feeding practices of 43 nine-month old Malawian infants and children enrolled in the iLiNS-DOSE clinical study (#NCT00945698) that took place in the Mangochi district of Malawi. The United States Department of Agriculture Nutrient Database (http://ndb.nal.usda.gov/ndb/search/list) was used estimate the macro- and micronutrient content of components of their diets. We then used linear programming to calculate combinations of ingredients that resembled the mean energy and nutrient values for diets identified from the dietary survey. **Table S3B** lists the composition of food ingredients in M8.

The M8 diet was prepared for administration to gnotobiotic animals as follows: Meseca® corn flour was obtained from Restaurant Depot (College Point, NY). The remaining ingredients were purchased from Whole Foods Supermarkets. The diet was prepared in batches of 20 kg. A relish containing 2kg mustard greens, 1.5kg onions and 1.5kg tomatoes was pureed in a food processor (Robot Coupe Model R23, Jackson, MS) and the puree was cooked in 1L of water for 60 minutes on a Corning stirrer/hot plot (high setting, until browned). After cooking, the relish was combined with a pureed mixture of 1kg ground peanuts, 700g soaked red kidney beans, 1kg canned pumpkin and 2.5kg peeled bananas in an industrial mixer (Globe SP30P 30-quart pizza mixer; gear speed 1; Globe Food Equipment Company, Dayton, OH). Corn flour

(5kg) and hot, freshly autoclaved water (5L) were then added slowly and mixed using the industrial mixer for 5 min.

Purified powdered S-BMO or inulin (from dahlia tubers, Santa Cruz Biotechnology, Inc., Product Number 9005-80-5) were then added to the M8 diet at the concentrations noted in the main text. Batches of unsupplemented M8, S-BMO-supplemented M8 or inulin-supplemented M8 diets were allowed to cool in clean plastic containers in a 4°C cold room prior to vacuum packing in 500 g aliquots in FDA/USDA-compliant poly-nylon vacuum pouches (#S-7556; Uline, Pleasant Prairie, WI). Sealed diet aliquots were double-bagged and sterilized by irradiation (20-50 kGy) within 24 hours of production (Steris Co; Chicago, IL).

For experimental convenience and consistency, dry pelleted versions of M8 and S-BMO-supplemented M8 were also obtained from Dyets, Inc. (Bethlehem, PA). Pellets were extruded with a 1/2-inch diameter for gnotobiotic mouse studies and a 3/16-inch diameter for gnotobiotic piglet studies.

The nutritional content of all cooked and irradiated custom diets was defined by N.P. Analytical Laboratories [St Louis, MO (**Table S3C**)]. Measurement of chow consumption by gnotobiotic animals confirmed comparable caloric and nutritional consumption between paste and pelleted versions of the diets (data not shown). Irradiated food was stored at 4°C for up to six months. Sterility was determined by resuspending a small aliquot of each batch of food in pre-reduced GMM under anaerobic conditions and incubating the suspension for 3 days at 37°C. Sterility was further verified by subculture on pre-reduced anaerobic GMM agar plates.

Studies involving gnotobiotic mice

All gnotobiotic mouse experiments were performed using protocols approved by the Washington University Animal Studies Committee. Male germ-free C57Bl/6J mice were maintained in sterile, flexible, plastic gnotobiotic isolators (Class Biologically Clean Ltd.,

Madison, WI) under a strict 12-hour cycle (lights on at 0600, off at 1800 h). Mice were fed an autoclaved low-fat, polysaccharide-rich chow (LF/HPP) diet (B&K University, East Yorkshire, U.K.; diet 7378000) from weaning until 3 days prior to the beginning of an experiment. At that time, 4.5-week old male animals were switched to the M8 diet or the M8 diet supplemented with S-BMO or inulin for the remainder of the experiment. Mice received a single oral gavage (200 μL) of the complete 25-member culture collection or subsets of the collection 3 days after switch to one of the M8-based diets.

Quantitative magnetic resonance (qMR) analysis of body composition - Body composition was defined using an EchoMRI-3in1 instrument (EchoMRI, Houston, TX). Each mouse was transported from its gnotobiotic isolator to the MR instrument in a HEPA filter-capped glass vessel.

Sample collection - Fecal samples were collected at defined times after gavage. At the time of euthanasia, blood, cecal contents, liver, brain and skeletal muscle (soleus or gastrocnemius) were collected and immediately frozen in liquid nitrogen. Soleus and gastrocnemius muscle samples were freeze-clamped after euthanasia *in situ* using pre-cooled forceps prior to harvest and storage.

Micro-computed tomography - Femurs were harvested from mice at time of euthanasia, cleaned of soft tissue, fixed for 24 hours in 10% formalin and stored at 4°C in 70% ethanol until scanning. Micro-computed tomography was performed using a μCT 40 desktop cone-beam instrument (ScanCO Medical, Brüttisellen, Switzerland). For cortical analyses, 200-300 slices were taken for each sample in the transverse plane with a 6 μm voxel size (high resolution). For all cortical scans, slices began at the midpoint of the femur and extended toward the distal femur. Boundaries of and thresholds for bone were drawn manually using μCT 40 software.

Calculation of volumetric parameters (bone volume/tissue volume, bone mineral density and cortical thickness) was automated with custom scripts.

Histological analysis of distal femur morphology – Previously fixed femurs were decalcified by incubation at room temperature in 15 mL of 14% EDTA, pH 7.2 (#EDS; Sigma Aldrich) for 14 days with gentle agitation. The EDTA solution was replaced daily. Decalcified femurs were subsequently washed four times with deionized water, rehydrated with stepwise immersions in 30% ethanol, 50% ethanol and 70% ethanol (30 minutes each), embedded in paraffin, cut into 5 μm thick longitudinal sections and stained for tartrate-resistant acid phosphatase (TRAP) for osteoclasts (Oddie et al., 2000) and hematoxylin for nuclei. Sections were imaged at 100x magnification with an Olympus U-CMAD3 microscope. The resulting images were analyzed using BioQuant OSTEO 2012 software (BIOQUANT Image Analysis Co.; Nashville, TN). Personnel were blinded to sample identifications. The region of interest for analysis of trabecular bone was defined for each section as the 2-dimensional area beginning 750 μm proximal to the distal growth plate and extending to the very proximal edge of the growth plate and spanning the full width of the bone. Osteoclasts were identified as TRAP+ cells. Regions located adjacent to the growth plate where osteoclast density was too high to accurately count cells were excluded from the analysis.

*Multiplex sequencing of amplicons generated from bacterial 16S rRNA gene*s - Genomic DNA was extracted by resuspending fecal pellets (~50 mg) or cecal contents (~50 mg) in a solution containing 500 μL of extraction buffer [200 mM Tris (pH 8.0), 200 mM NaCl, 20 mM EDTA], 210 μL of 20% SDS, 500 μL phenol:chloroform:isoamyl alcohol (pH 7.9, 25:24:1, Ambion) and 500 μL of 0.1-mm diameter zirconia/silica beads. Cells were mechanically disrupted using a bead beater (BioSpec Products, Bartlesville, OK; maximum setting for 4 min at room temperature),

followed by extraction with phenol:chloroform:isoamyl alcohol and precipitation with isopropanol.

Amplicons (~365bp) spanning variable region 4 (V4) of the bacterial 16S rRNA gene were generated by polymerase chain reaction (PCR) using primers and cycling conditions described in an earlier report (Subramanian et al., 2014). PCR primers incorporated sample-specific barcodes allowing samples to be subjected to multiplex sequencing using an Illumina MiSeq instrument (paired-end 250 nt reads). Paired V4-16S rRNA sequences were trimmed to 200 bp and merged into a single sequence with Flash software (Magoc and Salzberg, 2011). Merged sequences were filtered for low quality reads and binned according to their sample-specific barcodes. Reads were clustered into 97%ID OTUs using UCLUST (Edgar, 2010) and the Greengenes references OTU database. Reads that failed to hit the reference dataset were clustered *de novo* using UCLUST. A representative OTU set was created using the most abundant OTU from each bin. Reads were aligned using PyNAST (Caporaso et al., 2010). A custom dataset of manually curated NCBI bacterial taxonomy (Ridaura et al., 2013) was used to train the Ribosomal Database Project (RDP) version 2.4 classifier (Wang et al., 2007b) and to assign taxonomy to picked OTUs. Validation of this assignment strategy is described in a previous publication (Ridaura et al., 2013). For analysis of bacterial abundance and colonization efficiency, rare OTUs (<0.1% relative abundance on average across all samples collected from a given treatment group and timepoint) were removed.

Microbial RNA-Seq - Protocols for microbial RNA-Seq are described in our previous publications (Faith et al., 2011; McNulty et al., 2013; Turnbaugh et al., 2009). A slightly modified procedure was followed for this study. Cecal contents (100-300mg), collected 44 days after gavage, were suspended in 1 mL RNAprotect bacteria reagent (Qiagen), vortexed for 5 min at room temperature and centrifuged for 10 min at 5000 x g (4°C). After decanting the supernatant,

pelleted cells were suspended in 500 μL extraction buffer (200 mM NaCL, 20 mM EDTA), 210 μL of 20% SDS, 500 μL of phenol:chloroform:isoamyl alcohol (pH7.9, 125:24:1, Ambion) and 250 μL of acid-washed glass beads (Sigma-Aldrich, 212-300 μm diameter). Microbial cells were lysed by mechanical disruption using a bead beater (Biospec; maximum setting; 5 min at room temperature), followed by phenol:chloroform:isoamyl alcohol extraction and precipitation with isopropanol. RNA was treated with RNAse-free TURBO-DNAse (Ambion) and 5S rRNA and tRNAs were removed (MEGAClear columns, Ambion). A second DNAse treatment was performed (Baseline-ZERO DNAse; Epicenter). RNA integrity was checked by gel electrophoresis (100 ng, 1% agarose gel). 16S and 23S rRNA was depleted using the Ribo-zero meta-bacterial rRNA removal kit (Epicenter), followed by ethanol precipitation. cDNA was synthesized using SuperScript II (Invitrogen), followed by second strand synthesis with RNAseH, *E. coli* DNA polymerase (NEB) and *E. coli* DNA ligase (NEB). Samples were sheared using a BioRuptor XL sonicator (Diagenode); 200-300 bp fragments were gel selected and prepared for sequencing.

Multiplex microbial RNA-Seq for gnotobiotic mouse experiments was performed using an Illumina Hi-Seq2000 instrument to generate 10.78±1.45 million (mean±SD) unidirectional 101 nt reads per sample. Sequencing reads were split according to the 8-bp barcodes used to label each of the twelve samples pooled together per HiSeq lane (sequencing was performed over three lanes with reads pooled for analysis).

Reads were mapped to genes in a custom database of 25 unique draft genomes included in the defined Malawian infant bacterial community. If a read mapped to more than one location in a genome or to multiple genomes, the counts for each gene were added according to the gene's fraction of unique-match counts. Read counts (rounded to the nearest integer) for

each gene were split by genome and then normalized and analyzed with the R statistical package, DESeq (Anders and Huber, 2010).

Community profiling by sequencing (COPRO-Seq) – Protocols for determining bacterial community composition by COPRO-Seq are described in our previous publication (McNulty et al., 2011). Briefly, DNA samples were sheared using a BioRuptor Pico sonicator (Diagenode); 200-300 bp fragments were gel selected and prepared for sequencing. Samples were sequenced in multiplex using an Illumina NextSeq instrument to generate unidirectional 75 nt reads. Sequencing reads were split according to 8 bp barcodes assigned to each sample, processed and mapped to the draft genomes of members of the Malawian infant's culture collection using COPRO-Seq software (https://github.com/nmcnulty/COPRO-Seq).

Mass spectroscopy of metabolites contained in biospecimens obtained from gnotobiotic animals

Targeted ultra high performance liquid chromatography-mass spectrometry (UPLC-MS) of liver, muscle, brain, serum and cecal contents – Liver, brain, muscle and cecal samples were homogenized in 50% aqueous acetonitrile containing 0.3% formic acid (50 mg wet weight tissue/mL solution) using a high-speed homogenizer (IKA #EW-04739-21) set at maximum speed for 30-45 seconds. Samples were maintained on ice or dry ice throughout this procedure.

Amino acids, acylcarnitines, organic acids, acyl CoAs and ceramides were analyzed using stable isotope dilution techniques. Amino acids and acylcarnitine measurements were made by flow injection tandem mass spectrometry using sample preparation methods described previously (An et al., 2004, Ferrara et al., 2008). Data were acquired using a Waters Acquity™ UPLC system equipped with a TQ (triple quadrupole) detector and a data system controlled by MassLynx 4.1 operating system (Waters, Milford, MA). Organic acids were quantified according to a previously published protocol (Jensen et al., 2006) using Trace Ultra

GC coupled to ISQ MS operating under Xcalibur 2.2 (Thermo Fisher Scientific, Austin, TX). Acyl CoAs were extracted and purified as described (Deutsch et al., 1994; Magnes et al., 2005; Minkler et al., 2008), and analyzed by flow injection analysis using positive electrospray ionization on Xevo TQ-S, triple quadrupole mass spectrometer (Waters, Milford, MA). Heptadecanoyl CoA was employed as an internal standard. Ceramides were extracted using a previously reported procedure (Merrill et al., 2005) and analyzed by flow injection tandem mass spectrometry using a Xevo TQS spectrometer (Waters Milford, MA,) for precursors of m/z 264.

Non-targeted gas chromatography/mass spectrometry (GC-MS) of serum - Serum samples were spiked with perdeuterated myristic acid (D27-C14:0) as an internal standard for retention time locking (RTL IS). Following treatment with 7.5 volumes of methanol, the mixture was centrifuged and the supernatant was decanted and dried. Derivatization of all dried supernatants for GC-MS followed a method adapted with modifications from Roessner et al. (2000). Reagents were from Sigma-Aldrich (St. Louis, MO), unless otherwise noted. Briefly, certain reactive carbonyls were first methoximated at 50°C with a saturated solution of methoxyamine hydrochloride in dry pyridine, followed by replacement of exchangeable protons with trimethylsilyl (TMS) groups using *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide with a 1% v/v catalytic admixture of trimethylchlorosilane (Thermo-Fisher Scientific), also at 50°C. GC-MS methods generally used a 6890N GC connected to a 5975 inert single-quadrupole MS (Agilent, Santa Clara, CA). A large-volume ProSep inlet (Apex Technologies, Inc.; Independence, KY) enabled programmed temperature vaporization and diversion of heavy contaminants away from the GC and MS. The two wall-coated, open-tubular GC columns connected in series were both from J&W/Agilent (part 122-5512, DB5-MS; 15 m in length, 0.25 mm in diameter, with a 0.25 μm luminal film). Prior to each run, initial inlet pressures were empirically adjusted such that the resulting retention time of the TMS-D27-C14:0 standard was

set at ~16.727 min. Under these conditions, derivatized metabolites eluted from the column and reached the electron-ionization source in the MS at known times. A mid-column, microfluidic splitter (Agilent) provided a means for hot back-flushing of the upstream GC column at the end of each run while the oven was held at 325°C for a terminal "bake-out" [another antifouling and anti-carryover measure analogous to that described in (Chen et al., 2009)]. During this terminal "bake-out," the inlet was also held at 325°C and purged with a large flow of the carrier gas, helium. Positive ions generated with conventional EI at 70 eV were scanned broadly from 600 to 50 m/z in the detector throughout the run.

Raw data from Agilent's ChemStation software environment were imported into AMDIS (Automatic Mass Spectral Deconvolution and Identification Software; Stein, 1999). Deconvoluted spectra were identified, to the extent possible, using several commercial and public spectral libraries. Our primary source was the Fiehn GC-MS Metabolomics RTL Library (a gift from Agilent Technologies, Santa Clara, CA; part number G1676-90000). Additional spectra for comparison were gleaned from the Golm Metabolome Library (Kopka et al., 2005), the commercial NIST/EPA/NIH Mass Spectral Library and our own purpose-built spectral library. Where indicated, peak alignment was performed with SpectConnect freeware (Styczynski et al., 2007). Chemometrics were performed with Mass Profiler Professional (Agilent), along with our own custom macros, written in Visual Basic for use in the Excel software environment.

Non-targeted analysis of brain metabolites using GC-MS - Left cerebral hemispheres of mouse brains were homogenized using the same procedures described above for UPLC-MS of liver, serum and cecal contents. Homogenates were centrifuged (20,800 x g for 10 min at 4°C). A 200µL aliquot of the resulting supernatant was transferred to a clean tube and combined with 400 µL ice-cold methanol. The mixture was subsequently vortexed and centrifuged, and a 500

μL aliquot of the resulting supernatant, together with 10 μL of lysine-$^{13}C_6$,$^{15}N_2$ (2 mM), was evaporated to dryness using a speed vacuum. To derivatize the sample, 80 μL of a solution of methoxylamine (15 mg/mL in pyridine) was added to methoximate reactive carbonyls (incubation for 16 h for 37°C), followed by replacement of exchangeable protons with trimethylsilyl groups using *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (MSTFA) with a 1% (vol/vol) catalytic admixture of trimethylchlorosilane (Thermo-Fisher Scientific) (incubation for 1h at 70°C). Heptane (160 μL) was added and a 1 μL aliquot of each derivatized sample was injected into an Agilent 7890A gas chromatography system, coupled with a 5975C mass spectrometer detector (Agilent, CA). The quantity of each metabolite was determined from spike-in internal standards.

Nontargeted UPLC-MS of brain metabolites - Frozen mouse left cerebral hemisphere samples were homogenized with 20 vol/wt of cold methanol and 1 vol/wt cysteine $^{13}C_6$,$^{15}N_2$ (4 mM). Samples were subsequently incubated at -20°C for 1h and centrifuged 10 min at 20,800 x g. The resulting supernatant (300μL) was collected and dried in a SpeedVac at room temperature. Dried samples were resuspended in 100 μL of 95:5 water:ethanol, clarified for 5 min by centrifugation at 20,800 x g for 10 min at 4°C and the supernatant was separated for UPLC-MS. Analyses were performed on a Waters Acquity I Class UPLC system (Waters Corp., Milford, MA) coupled to an LTQ-Orbitrap Discovery (Thermo Fisher Corporation). Mobile phases used for negative ionization were (A) 5 mM ammonium acetate in water and (B) 5 mM ammonium acetate in acetonitrile/water (95%/5%).

Experiments involving gnotobiotic piglets

All experiments involving pigs were performed under the close supervision of a veterinarian using protocols approved by the Washington University Animal Studies Committee.

Derivation of germ free piglets -Pregnant domestic sows (mixture of Landrace and Yorkshire genetic backgrounds, mated to Duroc breed domestic boars) were obtained from Oak Hill Genetics (Ewing, IL). Piglets were delivered one day prior to the anticipated date of farrow. To prepare for surgery, the sow was sedated with ketamine (20 mg/kg administered intramuscularly) and anesthetized with isofluorane (2-3%, delivered by mask or endotracheal tube). A local incisional block was achieved using 60-80 mL of 2% lidocaine injected subcutaneously. The sow's paralumbar abdominal area was shaved, washed, and disinfected with povidone-iodine (1% titratable iodine). Piglets were delivered by making a paralumbar abdominal incision on the sow, exteriorizing the uterus and removing piglets individually into the sterile surgical field. Each piglet's umbilical cord was tied off to prevent bleeding and then severed. Piglets were then immediately passed, prior to their first breath, through a sterile tank filled with 10% betadine into a sterile gnotobiotic 'nursery' isolator made of flexible plastic. Piglets were revived within the 'nursery' isolator and kept on a heated pad until the procedure was complete. At the completion of surgery, the sow was euthanized by pentobarbital overdose (>150 mg/kg intravenously).

Immediately after piglets were revived (breathing independently and moving), they were transported within the nursery isolators from the operating room to a gnotobiotic facility. Within 24 hours, all piglets were transferred from the initial nursery isolators to larger, steel gnotobiotic isolator tubs (Class Biologically Clean; Madison, WI) where they were maintained for the duration of the experiment. Steel isolators were equipped with a flexible plastic canopy allowing trained personnel to examine and interact with the animals. Gnotobiotic piglets were group-housed (three piglets per isolator, complying with USDA animal housing regulations) and closely monitored throughout the experiment.

The room where gnotobiotic piglets were housed in their isolators was maintained at 85°F-90°F. Gnotobiotic isolators were maintained at 95°F-100°F using electric heaters positioned near their exterior (Cuori #HD904-A7Q 1500-watt electric oil-filled radiant portable heaters). Thermometer probes were positioned near inlets and outlets for airflow to monitor isolator temperatures. Piglet weights were measured using a hanging scale after carefully tying a rope around the piglet's abdomen and forelimbs.

Experimental design - Piglets were initially fed a powdered sow's milk replacement (Soweena Litter Life, Merrick #C30287N). The powdered sow's milk replacement was prepared in 120g vacuum-sealed, gamma-irradiated packets for reconstitution in gnotobiotic isolators (120g/ 1L autoclaved water). Piglets were fed by bottle for the first 2-3 days of life as they were trained to drink independently from troughs. Bottle-feeding was withdrawn when all piglets were observed to be drinking independently. Piglets were slowly weaned from sow's milk replacement formula to custom M8 pelleted chow starting at postnatal day 5 of life and completing on day 11. To do so, pelleted chow was introduced as creep feed in addition to formula, and the amount of formula provided to piglets was reduced stepwise until all piglets were observed eating pelleted chow. A small amount of sterile water was added to pelleted chow to encourage eating, and water was made available *ad libitum* to all animals in separate troughs. Behavior and the condition of all piglets were monitored multiple times each day to ensure their well-being. Piglets were fed at 3-hour intervals on days 1-2 after birth, at 4-hour intervals during days 3-4 and at 8-hour intervals from postnatal day 5 until the end of the experiment.

The germ-free status of piglets prior to colonization on postnatal day 3 was confirmed by aerobic and anaerobic culture of rectal swabs in LYHBHI medium (Sokol et al., 2008). Anaerobic cultures were taken by placing rectal swabs in anaerobic transport medium tubes

(Anaerobe Systems ATM 6.0ML #AS-911) for transport to an anaerobic chamber (atmosphere: 20% CO_2, 5% H_2, and 75% N_2).

Gnotobiotic piglets were colonized at day 3 of life by oral gavage. Briefly, equal volumes of overnight pre-reduced LYHBHI cultures of the 17 strains from the Malawian infant's bacterial culture collection that lacked more than three known virulence factor were mixed, diluted 1:10 in LYHBHI medium and mixed with an equal volume of pre-reduced PBS containing 30% glycerol [final glycerol concentration 15% (vol/vol)]. Aliquots of the bacterial consortium were stored at -80°C until use. For each piglet, 1 mL of this inoculum was resuspended in 10 mL of sow's milk replacement formula and introduced by intragastric gavage using a Kendall Kangaroo™ 2.7 mm diameter feeding tube (Cat #8888260406; Covidien, Minneapolis, MN).

Piglets were removed from the gnotobiotic isolator, sedated with ketamine (20 mg/kg IM), anesthetized with isofluorane and euthanized by pentobarbital overdose (>150 mg/kg IV). Whole blood was collected under anesthesia in Vacutainer tubes (BD #366404), immediately prior to euthanasia, stored on ice for at least 30 minutes and then centrifuged at 4000 × g for 5 minutes. Serum was collected, separated into small aliquots, and stored at -80°C. Other tissues were collected immediately after euthanasia, frozen in liquid nitrogen and stored at -80°C. Muscle samples were collected from the biceps femoris immediately after euthanasia and frozen in liquid nitrogen within 3-4 minutes.

Microbial RNA-Seq – RNA was extracted from fecal samples obtained at the time euthanasia and microbial RNA-Seq was performed using an Illumina NextSeq instrument [42.5±15.2 (mean±SD) million unidirectional 75 nt reads per sample]. Reads were demultiplexed using the 8 nt sample-specific barcode adapters and processed/mapped to the draft genomes of

strains in the culture collection using a set of custom scripts. Read counts were normalized and analyzed using DESeq (Anders and Huber, 2010).

Gut mucosa glycan analysis by triple quadrupole mass spectrometry – Mucosal samples were collected from three segments of the gut: (i) proximal small intestine (defined as a 10 cm segment distal to the stomach); (ii) distal small intestine (a 10 cm segment proximal to the cecum), and (iii) the spiral colon (10 cm segment distal to the cecum). Each of these segments was opened with clean steel scissors and luminal contents were gently removed with a clean steel spatula. The mucosa was gently scraped with a separate clean steel spatula and scrapings from each segment immediately frozen in liquid nitrogen. In addition, cecal contents and feces were recovered.

Each sample was diluted in water to 0.1 mg/10 µL, vortexed and briefly centrifuged at 2,000 x g. 100 µL of the resulting supernatant was taken and mixed with 100 µL of 0.2 M trifluoroacetic acid (TFA), then heated at 80°C for 1 hour. The solvent was then dried using a centrifugal evaporator. Samples were subsequently reconstituted in 200 µL of water and centrifuged at 21,130 x g for 30 min. 10 µL of the resulting supernatant was then added to 40 µL water and centrifuged at 21,130 x g for 20 min.

Analysis was performed on an Agilent ultra high-pressure liquid chromatography-electrospray ionization (UPLC-ESI)-triple quadrupole (QqQ) mass spectrometer in dynamic multiple reaction monitoring (MRM) mode. The Agilent 1290 infinity UPLC system incorporates a binary solvent system of solvent A [3% acetonitrile/water (v/v) in 0.1% formic acid (FA)] and solvent B [95% acetonitrile/water (v/v)] with an injection volume of 1 µL and flow rate of 0.3 mL/min with a Waters ACQUITY UPLC BEH amide column. Samples were run under isocratic conditions of 85% solvent A and 15% solvent B with a total run time of 3 min. The UPLC system is coupled to an Agilent 6490 QqQ mass spectrometer that was run in the

positive mode. The MRM transitions used for monitoring sialic acid were ($[M + 1H]^{1+}$ 310.2➔*m/z* 197.0 and *m/z* 121.1). MRM data were analyzed using Agilent MassHunter Quantitative Analysis B.05.02 software.

Measurement of serum leptin, insulin and triglycerides - Mouse serum triglycerides were measured using a DxC600 clinical analyzer (Beckman 445850; Brea, CA). Mouse insulin and leptin were measured using a duplex assay and an SI-2400 electrochemiluminescent imager (Meso Scale Discovery K15124C-1; Rockville, MD).

Porcine insulin and leptin were measured using a cross-reactive duplex assay developed for human sera (Meso Scale Discovery K151BYC-1; Rockville, MD). Porcine leptin was also measured with a human leptin kit (R&D Systems DLP00; Minneapolis, MN) as well as a porcine-specific kit from Biomatik (EKU05595; Wilmington, DE) using an SpectraMax M2e plate reader (Molecular Devices; Sunnyvale, CA). Porcine leptin was below the limit of detection by both methods.

***In vitro* S-BMO consumption and growth assays**

Frozen stock cultures were recovered by incubation overnight at 37°C in an anaerobic chamber in pre-reduced LYHBHI broth (Sokol et al., 2008). For incubation with S-BMO, strains were first grown overnight in 5 mL pre-reduced LYHBHI to an OD_{600} of approximately 0.8, then pelleted by centrifugation at 2600 x *g* for 5 minutes and washed twice with pre-reduced sterile PBS. Cell pellets were resuspended in 5 mL pre-reduced PBS containing 5% (wt/vol) and incubated at 37°C for 24 hours under anaerobic conditions (atmosphere: 20% CO_2, 5% H_2, and 75% N_2). Cells were pelleted by centrifugation at 2600 x *g* for 5 minutes, and the resulting culture supernatants were collected for UPLC-MS or GC-MS analysis of metabolites.

For *E. coli* growth experiments, starter cultures of *E. coli* strain MC1 were grown overnight in pre-reduced LYHBHI to an OD_{600} of approximately 0.8, pelleted, washed as above

and resuspended in 5 mL pre-reduced sterile PBS. *E. coli* cultures were started using a 100-fold dilution in a 200 μL volume of pre-reduced Davis Minimal Broth without dextrose (Sigma-Aldrich #15758, Saint Louis, MO) supplemented with 0.5% glucose, 0.5% sialic acid (Santa Cruz Biotechnology, Santa Cruz, CA), 2% S-BMO, 2% *B. fragilis* conditioned S-BMO (prepared by a 24 hour incubation of *B. fragilis* with S-BMO, as described above) or PBS only. All culture assays (*n*=8 technical replicates) were performed with negative controls (no *E. coli* inoculum) in a 96-well plate format. The OD_{600} of cultures was monitored over a 24-hour period using a Biotek Eon plate reader (Biotek Instruments, Winooski, VT) within the anaerobic Coy chamber.

Statistical analysis

Multivariate analyses (O-PLS-DA) of metabolomics data were performed using the R package, Muma (Gaude et al., 2013). Univariate analyses (t-tests and ANOVA) were performed using Microsoft Excel or Prism 6.0 (GraphPad Software, Inc.). Heat maps were generated using Gene-E software (http://www.broadinstitute.org/cancer/software/GENE-E/index.html). Two-way ANOVA of HMO and secretor status data from the LCNI-5 and iLiNS studies was performed by treating both factors (mother's secretor status and infant HAZ bin) as categorical variables. HMO abundance was treated as a continuous response variable. Spearman's rank correlation analyses were performed in the R environment using the test.cor function.

SUPPLEMENTAL REFERENCES

An, J., Muoio, D.M., Shiota, M., Fujimoto, Y., Cline, G.W., Shulman, G.I., Koves, T.R., Stevens, R., Millington, D., and Newgard, C.B. (2004). Hepatic expression of malonyl-CoA decarboxylase reverses muscle, liver and whole-animal insulin resistance. Nature Med *10*, 268-274.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., and Knight, R. (2010). PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics *26*, 266-267.

Chen, A., Kozak, D., Battersby, B.J., Forrest, R.M., Scholler, N., Urban, N., and Trau, M. (2009). Antifouling surface layers for improved signal-to-noise of particle-based immunoassays. Langmuir *25*, 13510-13515.

Delcher, A.L.P., A.; Carlton, J.; Salzberg, S.L. (2002). Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res *30*, 2478-2483.

Deutsch, J., Grange, E., Rapoport, S.I., and Purdon, D.A. (1994). Isolation and quantitation of long-chain acyl-coenzyme A esters in brain tissue by solid-phase extraction. Anal Biochem *220*, 321-323.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460-2461.

Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., Clemente, J.C., Knight, R., Heath, A.C., Leibel, R.L., *et al*. (2013). The long-term stability of the human gut microbiota. Science *341*, 1237439.

Faith, J.J., McNulty, N.P., Rey, F.E., and Gordon, J.I. (2011). Predicting a human gut microbiota's response to diet in gnotobiotic mice. Science *333*, 101-104.

Ferrara, CT, Wang P, Stevens, RD, Neto EC, Bain JR, Choi Y, Keller MP, Kendziorski CM, Yandell BS, Wenner BR, *et al.* (2008). Genetic networks of liver metabolism revealed by integration of metabolic and transcriptomic profiling. PLoS Genet *4,* e1000034.

Gaude, E., Chignola, F., Spiliotopoulos, D., Spitaleri, A., Ghitti, M., Garcia-Manteiga, J.M., Mari, S., and Musco, G. (2013). muma, An R Package for Metabolomics Univariate and Multivariate Statistical Analysis. Current Metabolomics *1*, 180-189.

Goodman, A.L, Faith, J.J., Reyes, A., Moore, A., Dantas, G., and Gordon, J.I. (2011). Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. Proc Natl Acad Sci USA *108*, 6252-6257.

Jensen, M.V., Joseph, J.W., Ilkayeva, O., Burgess, S., Lu, D., Ronnebaum, S.M., Odegaard, M., Becker, T.C., Sherry, A.D., and Newgard, C.B. (2006). Compensatory responses to pyruvate carboxylase suppression in islet beta-cells. Preservation of glucose-stimulated insulin secretion. J Biol Chem *281*, 22342-22351.

Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmuller, E., Dormann, P., Weckwerth, W., Gibon, Y., Stitt, M., *et al*. (2005). GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics *21*, 1635-1638.

Magnes, C., Sinner, F.M., Regittnig, W., and Pieber, T.R. (2005). LC/MS/MS method for quantitative determination of long-chain fatty acyl-CoAs. Anal Chem *77*, 2889-2894.

Magoc, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957-2963.

McNulty, N.P., Wu, M., Erickson, A.R., Pan, C., Erickson, B.K., Martens, E.C., Pudlo, N.A., Muegge, B.D., Henrissat, B., Hettich, R.L., *et al*. (2013). Effects of diet on resource utilization by a model human gut microbiota containing Bacteroides cellulosilyticus WH2, a symbiont with an extensive glycobiome. PLoS Biol *11*, e1001637.

Merrill, A.H., Jr., Sullards, M.C., Allegood, J.C., Kelly, S., and Wang, E. (2005). Sphingolipidomics: high-throughput, structure-specific, and quantitative analysis of sphingolipids by liquid chromatography tandem mass spectrometry. Methods *36*, 207-224.

Minkler, P.E., Kerner, J., Ingalls, S.T., and Hoppel, C.L. (2008). Novel isolation procedure for short-, medium-, and long-chain acyl-coenzyme A esters from tissue. Anal Biochem *376*, 275-276.

Newburg, D.S. and Neubauer, S.H. (1995). Handbook of milk composition, Chapter 4 (Carbohydrates in milk). Academic Press, Inc. 273-338.

Ninonuevo, M.R., Park, Y., Yin, H., Zhang, J., Ward, R.E., Clowers, B.H., German, J.B., Freeman, S.L., Killeen, K., Grimm, R., *et al*. (2006). A strategy for annotating the human milk glycome. J Agr Food Chem *54*, 7471-7480.

LoCascio, R.G., Desai, P., Sela, D.A., Weimer, B., and Mills, D.A. (2010). Broad conservation of milk utilization genes in Bifidobacterium longum subsp. infantis as revealed by comparative genomic hybridization. Appl Environ Micob *76*, 7373-7381.

Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J., Duncan, A.E., Kau, A.L., Griffin, N.W., Lombard, V., Henrissat, B., Bain, J.R., *et al*. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science *341*, 1241214.

Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N., and Willmitzer, L. (2000). Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. Plant *23*, 131-142.

Sela, D.A., Chapman, J., Adeuya, A., Kim, J.H., Chen, F., Whitehead, T.R., Lapidus, A., Rokhsar, D.S., Lebrilla, C.B., German, J.B., *et al*. (2008). The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. Proc Natl Acad Sci USA *105*, 18964-18969.

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L.G., Gratadoux, J.J., Blugeon, S., Bridonneau, C., Furet, J.P., Corthier, G., *et al*. (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci USA *105*, 16731-16736.

Stein, S.E. (1999). An integrated method for spectrum extraction and compound identification from GC/MS data. J Am Soc Mass Spec *10*, 770-781.

Styczynski, M.P., Moxley, J.F., Tong, L.V., Walther, J.L., Jensen, K.L., and Stephanopoulos, G.N. (2007). Systematic identification of conserved metabolites in GC/MS data for metabolomics and biomarker discovery. Anal Chem *79*, 966-973.

Totten, S.M., Zivkovic, A.M., Wu, S., Ngyuen, U., Freeman, S.L., Ruhaak, L.R., Darboe, M.K., German, J.B., Prentice, A.M., and Lebrilla, C.B. (2012). Comprehensive profiles of human milk

oligosaccharides yield highly sensitive and specific markers for determining secretor status in lactating mothers. J Proteome Res 11, 6124-6133.

Totten, S.M., Wu, L.D., Parker, E.A., Davis, J.C., Hua, S., Stroble, C., Ruhaak, L.R., Smilowitz, J.T., German, J.B., and Lebrilla, C.B. (2014). Rapid-throughput glycomics applied to human milk oligosaccharide profiling for large human studies. Anal Bio Chem *406*, 7925-7935.

Turnbaugh, P.J., Ridaura, V.K., Faith, J.J., Rey, F.E., Knight, R., and Gordon, J.I. (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci Trans Med *1*, 6ra14.

Wu, S., Tao, N., German, J.B., Grimm, R., and Lebrilla, C.B. (2010). Development of an annotated library of neutral human milk oligosaccharides. J Proteome Res *9*, 4138-4151.

Wu, S., Grimm, R., German, J.B., Lebrilla, C.B. (2011). Annotation and structural analysis of sialylated human milk oligosaccharides. J Proteome Res *10*, 856-868.