

Online supplemental materials

Anthropometric measurements

A standard, calibrated scale and wall-mounted stadiometer was used to measure weight and height, respectively. Body Mass Index (BMI) was calculated as weight (kg)/height (m)². Waist circumference was measured to the nearest 0.1 cm at the superior edge of the iliac crest using the Gulick II measuring tape. Three trials of the above measurements were completed, and the mean measurements were used for analyses. Participants are included with BMI \geq 97th percentile on age-and-sex-standardised WHO growth chart¹, and excluded if with impaired glucose tolerance or having taken antibiotics in the previous three months.

Tanner Stage Assessment

Participants completed a Pubertal Self-Assessment Tool to obtain Tanner Staging ².

Oral Glucose Tolerance Test, Metabolic and Plasma Lipids Assessment:

Blood draws were completed via an intravenous catheter in the antecubital fossa after anthropometrics measurements. All measurements were taken after an 8-14 hour overnight fast. Water was consumed up to 2 hours before the study visit. Baseline fasting samples were obtained (starting time: 0 minutes), followed by an Oral Glucose Tolerance Test (OGTT) where participants consumed a standard glucose-containing drink in 75g GlucoDex bottle (Teva Canada) with a dose of 1.75g/kg to a max of 75g. Venous blood samples were obtained at 30, 60, 90 and 120 minutes after the drink was consumed.

Collected blood samples were tested for (reference range indicated where available):
fasting glucose (3.9-6.0 mmol/L), fasting insulin (15-345 pmol/L), glycated hemoglobin

(HbA1c) (<6.0%), C-Reactive Protein (CRP) (0.1-1.7mg/L), cholesterol (<4.40 mmol/L), triglycerides (<1.5 mmol/L), high-density lipoprotein (HDL-C) (>1.2 mmol/L), low-density lipoprotein (LDL-C) (<2.6 mmol/L). Insulin and glucose levels taken from the OGTT were used to calculate the following measures of pancreatic β cell function: HOMA-IR (glucose (mmol) x insulin (uIU/ml)/22.5; ≤ 3.22 for Tanner stage 2-3, ≤ 2.91 for Tanner stage 4-5) = Fasting Insulin [uU/mL] x Fasting Glucose; Whole Body Insulin Sensitivity Index (WBISI = $10,000 / \sqrt{[(\text{Fasting Glucose} \times \text{Fasting Insulin}) \times (\text{mean glucose} \times \text{mean insulin})]}$), validated in adults and children³; and insulinogenic index (IGI = $\Delta [\text{uU/mL}] (0-30 \text{ min}) / \Delta \text{glucose} [\text{mmol/L}] (0-30 \text{ min})$) as measure of early insulin secretion and β cell function.

Remaining blood samples from clinical measurements were transported to the laboratory on ice and fractionated into PBMC and plasma using Ficoll-Hypaque Plus Density Gradient Media (Cytiva, USA) following manufacturer's instructions on the same day. 29 blood samples were processed on the same day within 5 hours of reception, while remaining samples from the cohort were stored at 4°C overnight and processed the following day. Recovered PBMC were counted using trypan blue and viably frozen at $1-10 \times 10^6$ cells/mL in 1mL of freezing media containing 10% DMSO in FBS, and stored in liquid nitrogen until analysis. Plasma fractions were divided into 1mL aliquots and stored at -80°C until analysis.

Quantification of plasma neutrophil-associated proteins, LPS-binding proteins, and fecal calprotectin

Soluble proteins in plasma or fecal samples were quantified using enzyme-linked immunosorbent assay (ELISA) through commercially available kits (Reagent Table).

Stool Sample Collection and Processing

Individual stool samples were available for 53 of the 56 participants. Stool samples were collected either during a clinic visit, or at home one day before a visit and transported to the hospital on ice and stored at -80°C. Frozen stool samples were separated into aliquots ~ 0.2g (wet weight) on dry ice for absolute bacterial quantification, fecal DNA extraction and water content measurement, respectively, and the remaining samples at - 80°C until analysis.

16S rRNA Gene Sequencing Analysis

Stool genomic DNA was extracted with the QIAamp Fast DNA Stool Mini Kit (QIAGEN, Canada) following manufacturer's instructions. Extracted DNA concentrations were quantified by a NanoDrop 2000C spectrophotometer (Thermo Fisher Scientific, USA) in lab and a Qubit Fluorometer 2.0 (Invitrogen, USA) at the Centre for Applied Genomics (TCAG) at the Hospital for Sick Children. DNA samples were diluted to 5ng/nL based on the fluorometric assay and sent for library preparation. Library preparation and sequencing were performed at the University of Colorado, Anschutz Medical Campus (Aurora, USA). The V3 and V4 hypervariable regions of the 16S rRNA gene were amplified with previously published F338/R806 primers and sequenced on a MiSeq platform (Illumina, San Diego, CA, USA)⁴. Overlapping paired-end reads were processed with a DADA2⁵ pipeline in R and unique ASVs were assigned a taxonomy based on SILVA v138.1 rRNA database ^{6,7} updated Mar 10, 2021 with a minimum 95% sequence overlap and 90% bootstrap confidence interval. The changes to phylum name advised by International Committee on Systematics of Prokaryotes⁸ are currently not available in SILVA ^{6,7}. Therefore, the taxa reported in this manuscript were assigned based previous prokaryotic nomenclature rules, with manual change to new phyla for display based on The List of Prokaryotic names with Standing in Nomenclature⁹. The first mention of a phylum in the manuscript shows both the

previous name and updated name to avoid confusion. 16S rRNA gene sequencing cannot resolve between *Escherichia coli* and *Shigella* due to high genetic similarity^{10,11}, and was assigned to both based on reference database. In main text, we refer to this taxon as *Escherichia coli*.

Fecal-derived Bio-reactor Culture

Fecal samples were inoculated and maintained in continuous-flow bioreactor cultures as previously published^{12,13}. Aliquots of fecal inoculum were normalized by absolute numbers of bacteria/dry g stool per group (high biomass: 2.9×10^{11} bacteria/aliquot; low biomass: 2.7×10^{10} bacteria/aliquot – based on group median values). Briefly, partially thawed fecal samples (2-8g wet weight) were transferred into sterile bags containing 10mL of sterile degassed chemostat media and homogenized by hand to form slurries. Each fecal slurry was used to inoculate a 500mL Multifors bioreactor system vessel (Infors AG, Bottmingen/Basel, Switzerland) already containing 490 mL sterile degassed chemostat media at 37°C, pH 7, and stirring at 50 rpm. Vessels were maintained under anaerobic, batch fermentation conditions for 24h before switching to continuous fermentation conditions (on day one). Bioreactor media was fed at a flow rate of 500 mL/day to ensure a 24h turnover of vessel contents. Vessels were maintained for 28 days, but analysis done on samples collected in the first 18 days.

CyTOF Immune Profiling of PBMC

Antibodies were purchased from various vendors (Reagent Table); most were metal tagged in-house using MaxPar X8 labelling kits (Standard BioTools). The optimal staining concentration for each antibody was pre-determined by evaluating a signal-to-noise metric over a 3-fold serial dilution series (range 1/30-1/2430) on healthy donor PBMC or other cell source containing

subsets known to be positive and negative for each marker. Frozen PBMC were quickly thawed in 37°C water bath and immediately added dropwise to 8mL of pre-warmed tissue culture medium (TCM: RPMI-1640 with GlutaMAX, 10% fetal bovine serum (FBS), 0.1 mM NEAA, 1 mM sodium pyruvate, 25 mM HEPES, and 2 mM L-glutamine) with 25 U/ml Benzonase B (Key Resource Table). Cell samples were added 1 mL of TCMB, centrifuged at 300g for 8min, resuspended in 5mL of TCMB and centrifuged again at 300g for 5min. Cell pellets were resuspended in TCMB at $1-2 \times 10^6$ cells per 4mL of sample and incubated for 30min at 37 °C. Samples were then washed by adding 1-2mL of cell staining media (SM: phosphate buffered saline (PBS) + 1% bovine serum albumin (BSA)) and centrifuged at 300g for 5min. Cell pellets were resuspended in 25µL Fc-receptor block (Reagent Table) diluted to 1/20 in SM and pre-incubated at room temperature for 10 minutes prior to adding 25µL of the metal-tagged antibody cocktail. After an additional 30min at room temperature, cells were pelleted at 300g for 5min at 4°C, and washed twice with 2mL PBS at 300g for 5min at 4 °C before staining with 1µM Cisplatin for 5min at room temperature. Cisplatin was quenched with 2mL SM. Samples were underlay with calf serum, pelleted for 5 min at 300g, aspirated, and flicked to re-suspend. Cellular DNA was then labelled with 1 mL of 100 nM ^{191/193} Iridium in PBS containing 0.3% saponin and 1.6% formaldehyde overnight at 4 °C. On the following day, samples were washed in 4 volumes SM and pelleted at 800 x g at 4° C for 5 min. Cells were washed twice more with PBS and pelleted under the same conditions before re-suspending in Maxpar Cell Acquisition Solution (Reagent Table) at $2-5 \times 10^5$ /ml containing four-element EQ normalization beads (Reagent Table) according to the manufacturer's instructions. Personnel at the SickKids' Centre for Advanced Single Cell Analysis (CASCA) analyzed the samples on a Helios3 instrument equipped with a wide-bore injector at ~300 cells/second according to Standard BioTools'

protocols. The Helios software (v7.0.8493) generated integrated mass data files that were normalized and converted to FCS 3.0 datafiles for downstream analysis.

Gut microbiome data pre-processing and analyses

All analyses in the manuscript, unless indicated otherwise, were completed in R v3.6.2 and v4.1.2¹⁴. Filtering and visualization of the ASVs at different taxonomic levels, as well as calculation of alpha diversity were completed using phyloseq v1.30.0¹⁵. Overall, 2749 unique ASVs were detected when accounting for sequenced amplicon error using R package dada2 v1.14.0⁵. After applying filters of > 0.1 prevalence and > 0.0005 of mean relative abundance, 199 ASVs were used as the core microbiome for distance-based Redundancy Analysis¹⁶, completed using vegan v2.5-6¹⁷ and microViz v 0.10.5¹⁸, and hierarchical clustering, completed using ComplexHeatmap v2.2.0¹⁹ and dendextend v1.13.4²⁰. The Bray-Curtis distance was calculated using vegan v2.5-6¹⁷. The permutational test for variable association with gut microbiome composition was completed using adonis2() from vegan v2.5-6¹⁷, setting *by* = "margin" to perform separate significance test for each marginal term in a model with all other terms.

Absolute quantification of gut bacteria and stool water content

Frozen fecal samples were aliquoted, weighted, and normalized to 0.2g of wet feces per mL of 0.22µm-filtered Bacteria Staining Media (BSM, 1 mM EDTA, 0.05% Tween-20 in sterile PBS). After incubating for 15-30 minutes at room temperature, samples were homogenized by vortexing (1-5 min). To remove debris and large cells, samples were centrifuged at 700g for 1 min twice, and the fecal supernatant was collected. Next, 100ul of the fecal supernatant (equivalent to 20ug of wet stool) was washed with 1mL of BSM and centrifuged at 6000g for 4

min to recover bacterial cells. Pellets were washed once with 1mL of staining buffer and resuspended in the original volume (100ul). Bacterial cell suspensions were diluted 1/200 and 1/2000 in the staining buffer, and the diluted suspension was stained with Thiazole Orange (final concentration 420nM ; Anaspec) and Propidium Iodide (final concentration 20uM ; Sigma-Aldrich) for 15 min at room temperature. A known concentration of counting beads ($\sim 10^5$ beads, AccuCount beads, Spherotech) was added to the stained cells, and the total number of nucleic acid-positive cells was counted by flow cytometry on an LSRFortessa flow cytometer (BD Biosciences, SickKids Flow cytometry facility) according to the manufacturer's instructions. Bacteria enumeration by flow cytometry was acquired in the BD FACSDiva™ software. A threshold value of 400-800 was applied on the side scatter V405 channel, with the stopping gate set to 6000 beads.

Stool water content was determined from frozen stool aliquots (0.1-0.2 g) as the ratio of mass loss (wet weight/dry weight) after air drying over 48h at room temperature in a Hamilton SafeAire II fume hood (Thermo Scientific). The wet/dry weight ratio represents the stool water content of each sample. We used the calculated wet/dry weight ratio to normalize the microbial cell counts assessed by flow cytometry (in 20ug of wet stool) to bacterial cell counts in one gram of dry stool. The absolute abundance of each ASV identified by 16S rRNA gene sequencing was determined by multiplying the relative abundance of each ASV by the total bacterial biomass normalized by one gram of dry stool.

CytoTOF Data Analysis

FCS 3.0 datafiles were uploaded to CytoBank (Enterprise v10.4, Beckman Coulter Life Sciences) where all pre-processing steps were implemented. Data were pre-processed by arcsinh

transformation (cofactor = 5), with the minimum and maximum values for some markers being further adjusted to improve visualization. Standard pre-gating steps were then used to identify cell events as (^{140}Ce EQ $^-$ DNA1 $^{\text{all}}$), exclude aggregates and 195Cisplatin $^{\text{hi}}$ dead cells²¹. This population was further cleaned by removing CD3 $^+$ CD14 $^+$, CD3 $^+$ CD19 $^+$, CD19 $^+$ CD7 $^+$, or CD19 $^+$ CD14 $^+$ aggregates. Within the resulting live single cell “clean” population, low-density (LD) CD45 $^{\text{lo}}$ CD66b $^+$ neutrophils were distinguished from the CD45 $^{\text{hi}}$ CD66b $^-$ population which was further gated to identify the following subsets: TCR $\alpha\beta$ (CD3 $^+$ TCRgd $^-$), TCR $\gamma\delta$ (CD3 $^+$ TCRgd $^+$), B cells were identified as CD3 $^-$ CD14 $^-$ CD19 $^+$ cells, natural killer cells (NK)+innate lymphoid cells (ILC) were identified as CD3 $^-$ CD19 $^-$ CD14 $^{\text{all}}$ HLA-DR $^-$ CD7 $^+$ cells, and myeloid cells were identified as CD3 $^-$ CD19 $^-$ CD14 $^{\text{all}}$ HLA-DR $^+$ CD7 $^-$ cells. Among TCRab cells, different lineages were identified as CD4 $^+$, CD8a $^+$, double negative (CD4 $^-$ CD8a $^-$), and mucosal-associated invariant T cells (MAIT) cells (TCRVa7.2 $^+$ CD161 $^+$), with CD4 and CD8 separated into naïve (CD45RO $^-$ CD7 $^{\text{all}}$) and memory (CD45RO $^+$ CD7 $^{\text{all}}$) cells. B cells were separated based on receptor isotype into unswitched IgM $^{\text{hi}}$ B cells, unswitched IgD $^{\text{hi}}$ B cells, and switched IgM $^-$ IgD $^-$ B cells. Among myeloid cells, plasmacytoid dendritic cells (pDC) were identified as CD11c $^-$ CD14 $^-$ CD123 $^+$ CD38 $^+$ cells, and HLA-DR $^+$ CD11c $^+$ population is further separated into classical monocytes (CD14 $^+$ CD16 $^-$), non-classical monocytes (CD14 $^-$ CD16 $^+$), and classical dendritic cells (CD11b $^-$ CD35 $^-$ HLA-DR $^{\text{hi}}$ CD38 $^+$), of which conventional dendritic cells 1(cDC1) are marked by CD141, and conventional dendritic cells 2 (cDC2) are marked by CD1c. NK/ILC subsets were defined based on CD56 and CD16. The frequencies of all cell subsets in low and high biomass group were compared by permutation test using *compare.2.vectors()* function from *afex* package²², visualized in dot plots, and corrected for multiple testing using *p.adjust()* in *stats* package¹⁴. We performed UMAP²³ dimensionality reduction in CytoBank by equal

downsampling to 46213 events/sample (set to 10 neighbors, min distance 0.4). One sample was excluded due to low cell count (n = 23224).

Handling Missing Data and Normalization

Missing data in clinical measurements was imputed using multivariate imputation by chained equations (MICE) v3.16.0²⁴. Continuous variables were transformed into a standard normal distribution based on center, scale, and Yeo-Johnson transformation using caret v6.0-85²⁵.

Statistical analysis

Unless otherwise specified, the associations between fecal bacteria counts, taxa abundances, cardiometabolic measures, plasma neutrophil proteins, and other variables were evaluated by linear regression using `lm()` in the base R stats package¹⁴. Permutation tests were used to evaluate differentially abundant ASVs and immune cell populations with `afex` v1.3-0²². Multiple testing correction was completed using `p.adjust()` in stats package¹⁴. Multinomial logistic regression from `nnet` v7.3-12 was used to predict probability of cluster assignment based on normalized fecal bacteria counts²⁶. Gamma generalized linear regression with log-link function using `glm()` in stats was used to evaluate the association between *Escherichia* abundance and fecal and plasma calprotectin. ESEM of the relationships between three latent variables -- the gut microbiota, inflammation, and metabolic variation -- and various observed variables was completed in `lavaan`²⁷. All regressions were adjusted for age, sex and BMI. All visualization and analysis are completed with `tidyverse` v1.3.0²⁸.

Reagent table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-CD45 ⁸⁹ Y (clone HI30)	Standard BioTools	Cat#3089003B; RRID: AB_2938863

Mouse monoclonal anti-CD38 ¹¹⁴ Cd (clone HB7)	Biolegend, conjugated in house with ¹¹⁴ Cd	MaxPar ready custom formulation of Cat# 356602; RRID: AB_2561794
Mouse monoclonal anti-CD11c ¹¹⁵ In (clone Bu15)	Biolegend, conjugated in house with ¹¹⁵ In	Cat# 337221; RRID: AB_2562834
Mouse monoclonal anti-HLA-DR ¹⁴¹ Pr (clone L243)	Biolegend, conjugated in house with ¹⁴¹ Pr	Cat# 307651; RRID: AB_2562826
Mouse monoclonal anti-IgD ¹⁴² Nd (clone IA6-2)	Biolegend, conjugated in house with ¹⁴² Nd	Cat# 348235; RRID: AB_2563775
Mouse monoclonal anti-CD27 ¹⁴⁴ Nd (clone L128)	BD Biosciences, conjugated in house with ¹⁴⁴ Nd	Custom order of purified anti-human CD27 from BD Biosciences
Mouse monoclonal anti-CD8a ¹⁴⁶ Nd (clone RPA-T8)	Biolegend, conjugated in house ¹⁴⁶ Nd	Cat# 301002; RRID: AB_314120
Mouse monoclonal anti-CD35 ¹⁴⁷ Sm (clone E11)	Biolegend, conjugated in house with ¹⁴⁷ Sm	Cat# 333402; RRID: AB_1089032
Mouse monoclonal anti-CD4 ¹⁴⁸ Nd (clone SK3)	Biolegend, conjugated in house with ¹⁴⁸ Nd	Cat# 344625; RRID: AB_2563749
Mouse monoclonal anti-CD24 ¹⁵⁰ Nd (clone ML5)	Biolegend, conjugated in house with ¹⁵⁰ Nd	Cat# 311127; RRID: AB_2563733
Mouse monoclonal anti-CD123 ¹⁵¹ Eu (clone 6H6)	Biolegend, conjugated in house with ¹⁵¹ Eu	Cat# 306027; RRID: AB_2562823
Mouse monoclonal anti-CD7 ¹⁵² Sm (clone CD7-6B7)	Biolegend, conjugated in house with ¹⁵² Sm	Cat# 343111; RRID: AB_2563761
Mouse monoclonal anti-CD45RA ¹⁵³ Eu (clone HI100)	Biolegend, conjugated in house with ¹⁵³ Eu	Cat# 304102; RRID: AB_314406
Mouse monoclonal anti-CD161 ¹⁵⁵ Gd (clone HP-3G10)	Biolegend, conjugated in house with ¹⁵⁵ Gd	Cat# 339919; RRID: AB_2562836
Mouse monoclonal anti-CD138 ¹⁵⁸ Gd (clone MI15)	Biolegend, conjugated in house with ¹⁵⁸ Gd	MaxPar ready custom formulation of Cat# 356502; RRID: AB_2561790
Mouse monoclonal anti-IgM ¹⁵⁹ Tb (clone MHM-88)	Biolegend, conjugated in house with ¹⁵⁹ Tb	Cat# 314527; RRID: AB_2563776
Mouse monoclonal anti-CD14 ¹⁶⁰ Gd (clone M5E2)	Biolegend, conjugated in house with ¹⁶⁰ Gd	Cat# 301802; RRID: AB_314184
Mouse monoclonal anti-CD25 ¹⁶¹ Dy (clone 2A3)	BD Biosciences, conjugated in house with ¹⁶¹ Dy	MaxPar ready custom formulation of Cat# 347640; RRID: AB_400333
Mouse monoclonal anti-CD1c ¹⁶² Dy (clone L161)	Biolegend, conjugated in house with ¹⁶² Dy	MaxPar ready custom formulation of Cat# 331502; RRID: AB_1088995
Mouse monoclonal anti-CD127 ¹⁶³ Dy (clone eBioRDR5)	eBioscience, conjugated in house with ¹⁶³ Dy	Cat# 14-1278-82; RRID: AB_657591
Mouse monoclonal anti-CD279 ¹⁶⁴ Dy (clone EH12.1)	BD Biosciences, conjugated in house with ¹⁶⁴ Dy	Cat# 562138; RRID: AB_10897007
Mouse monoclonal anti-CD16 ¹⁶⁵ Ho (clone 3G8)	Biolegend, conjugated in house with ¹⁶⁵ Ho	Cat# 302002; RRID: AB_314202

Rat monoclonal anti-CXCR5 ¹⁶⁶ Er (clone RF8B2)	BD Biosciences, conjugated in house with ¹⁶⁶ Er	Cat# 552032, RRID: AB_394324
Mouse monoclonal anti-TCR-Va7.2 ¹⁶⁷ Er (clone 3C10)	Biolegend, conjugated in house with ¹⁶⁷ Er	Cat# 351702, RRID: AB_10900258
Mouse monoclonal anti-CD19 ¹⁶⁸ Er (clone HIB19)	Biolegend, conjugated in house with ¹⁶⁸ Er	Cat# 302247, RRID: AB_2562815
Mouse monoclonal anti-CD66b ¹⁶⁹ Tm (clone G10F5)	BD Biosciences, conjugated in house with ¹⁶⁹ Tm	MaxPar ready custom formulation of Cat# 555723; RRID: AB_39606
Mouse monoclonal anti-CD3 ¹⁷⁰ Er (clone UCHT1)	eBioscience, conjugated in house with ¹⁷⁰ Er	Cat# 16-0038-85, RRID: AB_468857
Mouse monoclonal anti-CD141 ¹⁷¹ Yb (clone 1A4)	BD Biosciences, conjugated in house with ¹⁷¹ Yb	MaxPar ready custom formulation of Cat# 559780; RRID: AB_397321
Mouse monoclonal anti-TCRgd ¹⁷² Yb (clone 5A6.E9)	Purified from hybridoma in house	
Mouse monoclonal anti-CD56 ¹⁷³ Yb (clone REA196)	Miltenyi Biotec, conjugated in house with ¹⁷³ Yb	Cat# 130-108-016, RRID: AB_2658728
Mouse monoclonal anti-CD45RO ¹⁷⁴ Yb (clone UCHL1)	Biolegend, conjugated in house with ¹⁷⁴ Yb	Cat# 304202, RRID: AB_314418
Mouse monoclonal anti-CD163 ¹⁷⁵ Lu (clone GHI/61)	Biolegend, conjugated in house with ¹⁷⁵ Lu	MaxPar ready custom formulation of Cat# 333602, RRID: AB_1088991
Mouse monoclonal anti-CCR7 ¹⁷⁶ Yb (clone G043H7)	Biolegend, conjugated in house with ¹⁷⁶ Yb	Cat# 353237, RRID: AB_2563726
Mouse monoclonal anti-CD11b ²⁰⁹ Bi (clone ICRF44)	Standard BioTools, conjugated in house with ²⁰⁹ Bi	Cat# 3209003B, RRID: AB_2687654
Cisplatin 1g	BioVision Research Products	1550-1000
Cell-ID™ Intercalator-Ir (500uM)	Standard BioTools	201192B
Biological samples		
Stool	Hospital for Sick Children	NCT04031222
PBMC	Hospital for Sick Children	NCT04031222
Plasma	Hospital for Sick Children	NCT04031222
Chemicals, peptides, and recombinant proteins		
Ficoll-Paque PLUS, 6x100ml/pack	Cytiva	CA95021-205L
Thiozole Orange, 10 mM in DMSO, UltraPure Grade	Anaspec	AS-83228
Propidium iodide, ≥95% HPLC, 100mg	Sigma-Aldrich	P4170
Trypan Blue solution	Sigma-Aldrich	T8154
DMSO, Hybri-Max™, sterile-filtered, ≥ 99.7%, 100mL	Sigma-Aldrich	D2650
RPMI-1640 with GlutaMAX	Life Technologies	61870-127
Fetal Bovine Serum	Wisent	080-150
MEM Non-Essential Amino Acids Solution (100X)	Life Technologies	11140-050

Sodium Pyruvate (100 mM)	Life Technologies	11360-070
HEPES, Free Acid 100g	Wisent	600-032 EG
L-Glutamine (200mM)	Life Technologies	25030-081
Benzonase® Nuclease, UltraPure	Sigma-Aldrich	E1014-25KU
Phosphate-buffered saline (PBS), Ca ²⁺ /Mg ²⁺ free, pH 7.2	Gibco	20012-050
Bovine serum albumin (BSA), Cohn Fraction V, ≥98%	Sigma-Aldrich	A7906
Human TruStain FcX™	BioLegend	422302
Saponin	Sigma-Aldrich	S7900
Formaldehyde, 16%, methanol free, UltraPure	Polysciences	18814-20
EDTA (0.5 M), pH 8.0, RNase-free	Invitrogen	AM9261
TWEEN® 20	Sigma-Aldrich	P5927
Maxpar Cell Acquisition Solution—200 mL	Standard BioTools	201241
EQ™ Four Element Calibration Beads—100 mL	Standard BioTools	201078
AccuCount Blank Particles	Spherotech	ACBP-50-10
Critical commercial assays		
QIAamp DNA Stool Mini Kit	QIAGEN	51504
MiSeq Reagent Kit v3 (600-cycle)	Illumina	MS-102-3003
Maxpar® X8 Antibody Labeling Kits	Standard BioTools	201142
Cell Viability Kit	BD Biosciences	349483
Human S100A8/S100A9 Heterodimer DuoSet ELISA	R&D Systems	DY8226-05
Human Neutrophil Elastase/ELA2 DuoSet ELISA	R&D Systems	DY9167-05
Human Myeloperoxidase DuoSet ELISA	R&D Systems	DY3174
Human CD14 DuoSet ELISA	R&D Systems	DY383-05
Human LBP DuoSet ELISA	R&D Systems	DY870-05
DuoSet ELISA Ancillary Reagent Kit 2	R&D Systems	DY008B
Deposited data		
SILVA v138.1 rRNA database	Quast et al. ⁶ and McLaren et al. ⁷	10.5281/zenodo.4587954
Primers		
16S rRNA-Forward Primer 338F: CCTACGGGAGGCAGCAG	Hara et al. ⁴	10.4049/jimmunol.1201257
16S rRNA-Reverse Primer 806R: CTACHVGGGTWTCTAAT	Hara et al. ⁴	10.4049/jimmunol.1201257
Software and algorithms		
R and stats v 3.6.2 and 4.1.2	R Core Team ¹⁴	https://www.R-project.org
dada2 v 1.14.0	Callahan et al. ⁵	https://benjjneb.github.io/dada2/index.html
FACSDiva v8.0.1	BD Biosciences	

Cytobank Enterprise version	Beckman Coulter	https://www.mybeckman.ca/flow-cytometry/software/cytobank-enterprise
phyloseq v 1.30.0	Paul J. McMurdie and Susan Holmes ¹⁵	http://dx.plos.org/10.1371/journal.pone.0061217
microViz v 0.10.5	David J.M. Barnett and Ilja C.W. Arts and John Pender ¹⁸	https://doi.org/10.21105/joss.03201
tidyverse v 1.3.0	Wickham et al. ²⁸	https://doi.org/10.21105/joss.01686
ComplexHeatmap v 2.2.0	Zuguang Gu and Roland Eils and Matthias Schlesner ¹⁹	10.1093/bioinformatics/btw313
dendextend v 1.13.4	Tal Galili ²⁰	10.1093/bioinformatics/btv428
vegan v 2.5-6	Dixon ¹⁷	https://CRAN.R-project.org/package=vegan
mice v3.16.0	Stef van Buuren and Karin Groothuis-Oudshoorn ²⁴	10.18637/jss.v045.i03
caret v6.0-85	Max Kuhn ²⁵	https://CRAN.R-project.org/package=caret
afex v 1.3-0	Singmann et al. ²²	https://CRAN.R-project.org/package=afex
nnet v 7.3-12	W. N. Venables and B. D. Ripley ²⁶	http://www.stats.ox.ac.uk/pub/MASS4
lavaan v 0.6-17	Yves Rosseel ²⁷	10.18637/jss.v048.i02

Supplemental references

1. World Health Organization. Growth reference 5-19 years - BMI-for-age (5-19 years). 2007. <https://www.who.int/tools/growth-reference-data-for-5to19-years/indicators/bmi-for-age>
2. World Health Organization. ANNEX H, SEXUAL MATURITY RATING (TANNER STAGING) IN ADOLESCENTS. In: Antiretroviral Therapy for HIV Infection in Infants and Children: Towards Universal Access: Recommendations for a Public Health Approach: 2010 Revision. Geneva: World Health Organization; 2010. (WHO Guidelines Approved by the Guidelines Review Committee). <https://www.ncbi.nlm.nih.gov/books/NBK138588/>
3. Retnakaran R, Qi Y, Goran MI, Hamilton JK. Evaluation of proposed oral disposition index measures in relation to the actual disposition index. *Diabetic Medicine*. 2009;26(12):1198–1203. doi:10.1111/j.1464-5491.2009.02841.x
4. Hara N, Alkanani AK, Ir D, Robertson CE, Wagner BD, Frank DN, Zipris D. Prevention of Virus-Induced Type 1 Diabetes with Antibiotic Therapy. *The Journal of Immunology*. 2012;189(8):3805–3814. doi:10.4049/jimmunol.1201257
5. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*. 2016;13(7):581–583. doi:10.1038/nmeth.3869
6. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*. 2012;41(D1):D590–D596. doi:10.1093/nar/gks1219
7. McLaren MR, Callahan BJ. Silva 138.1 prokaryotic SSU taxonomic training data formatted for DADA2. 2021 [accessed 2024 May 28]. <https://zenodo.org/record/4587955>. doi:10.5281/ZENODO.4587955
8. Oren A, Garrity GM. Valid publication of the names of forty-two phyla of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*. 2021 [accessed 2024 Aug 1];71(10). <https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.005056>. doi:10.1099/ijsem.0.005056
9. Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology*. 2020;70(11):5607–5612. doi:10.1099/ijsem.0.004332
10. Chattaway MA, Schaefer U, Tewolde R, Dallman TJ, Jenkins C. Identification of *Escherichia coli* and *Shigella* Species from Whole-Genome Sequences Ledebauer NA, editor. *Journal of Clinical Microbiology*. 2017;55(2):616–623. doi:10.1128/JCM.01790-16

11. Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Inbanathan FY, Veeraraghavan B. Accurate differentiation of *Escherichia coli* and *Shigella* serogroups: challenges and strategies. *New Microbes and New Infections*. 2018;21:58–62. doi:10.1016/j.nmni.2017.09.003
12. Gianetto-Hill CM, Vancuren SJ, Daisley B, Renwick S, Wilde J, Schroeter K, Daigneault MC, Allen-Vercoe E. The Robogut: A Bioreactor Model of the Human Colon for Evaluation of Gut Microbial Community Ecology and Function. *Current Protocols*. 2023;3(4):e737. doi:10.1002/cpz1.737
13. Granato A, Renwick S, Yau C, Kong T, Daigneault MC, Knip M, DIABIMMUNE Study Group, Allen-Vercoe E, Danska JS. Analysis of early childhood intestinal microbial dynamics in a continuous-flow bioreactor. *Microbiome*. 2024;12(1):255. doi:10.1186/s40168-024-01976-w
14. R Core Team. R: A language and environment for statistical computing. 2019. <https://www.R-project.org/>
15. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data Watson M, editor. *PLoS ONE*. 2013;8(4):e61217. doi:10.1371/journal.pone.0061217
16. Legendre P, Anderson MJ. Distance-Based Redundancy Analysis: Testing Multispecies Responses in Multifactorial Ecological Experiments. *Ecological Monographs*. 1999;69(1):1–24. doi:10.1890/0012-9615(1999)069[0001:DBRATM]2.0.CO;2
17. Dixon P. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*. 2003;14(6):927–930. doi:10.1111/j.1654-1103.2003.tb02228.x
18. Barnett D, Arts I, Penders J. microViz: an R package for microbiome data visualization and statistics. *Journal of Open Source Software*. 2021;6(63):3201. doi:10.21105/joss.03201
19. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32(18):2847–2849. doi:10.1093/bioinformatics/btw313
20. Galili T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*. 2015;31(22):3718–3720. doi:10.1093/bioinformatics/btv428
21. Bagwell CB, Inokuma M, Hunsberger B, Herbert D, Bray C, Hill B, Stelzer G, Li S, Kollipara A, Ornatsky O, et al. Automated Data Cleanup for Mass Cytometry. *Cytometry Part A*. 2020;97(2):184–198. doi:10.1002/cyto.a.23926
22. Singmann H, Bolker B, Westfall J, Aust F, Ben-Shachar MS. afex: Analysis of Factorial Experiments. 2023. <https://CRAN.R-project.org/package=afex>

23. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. 2018 [accessed 2024 Jun 1]. <https://arxiv.org/abs/1802.03426>. doi:10.48550/ARXIV.1802.03426

24. Buuren SV, Groothuis-Oudshoorn K. mice : Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. 2011 [accessed 2024 May 28];45(3). <http://www.jstatsoft.org/v45/i03/>. doi:10.18637/jss.v045.i03

25. Kuhn M. caret: Classification and Regression Training. *Astrophysics Source Code Library*. 2015 May 1:ascl:1505.003.

26. Venables WN, Ripley BD. *Modern applied statistics with S*. 4th ed. New York: Springer; 2002. (Statistics and computing).

27. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*. 2012 [accessed 2024 May 28];48(2). <http://www.jstatsoft.org/v48/i02/>. doi:10.18637/jss.v048.i02

28. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Golemund G, Hayes A, Henry L, Hester J, et al. Welcome to the Tidyverse. *Journal of Open Source Software*. 2019;4(43):1686. doi:10.21105/joss.01686