Differences in the Gut Microbiomes of Distinct Ethnicities Within the Same Geographic Area Are Linked to Host Metabolic Health

Qi Yan Ang

University of California San Francisco

Diana L Alba

University of California San Francisco

Vaibhav Upadhyay

University of California San Francisco

Jordan E. Bisanz

University of California San Francisco

Jingwei Cai

Pennsylvania State University

Ho Lim Lee

University of California San Francisco

Eliseo Barajas

University of California San Francisco

Grace Wei

University of California San Francisco

Andrew D. Patterson

Pennsylvania State University

Suneil K. Koliwad

University of California San Francisco

Peter James Turnbaugh (✉ peter.turnbaugh@ucsf.edu)

University of California San Francisco <https://orcid.org/0000-0002-0888-2875>

Research

Keywords: human gut microbiome, ethnicity, multi-omics, metabolic syndrome, obesity, biogeography

Posted Date: November 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-100084/v1>

Abstract

Background

The human gut microbiota exhibits marked variation around the world, which has been attributed to dietary intake and other environmental factors. However, the degree to which ethnicity-associated differences in gut microbial community structure and function are maintained following immigration or in the context of metabolic disease is poorly understood.

Results

We conducted a multi-omic study of 46 lean and obese East Asian (EA) and White (W) participants living in the San Francisco Bay Area. 16S rRNA gene sequencing revealed significant differences between ethnic groups in bacterial richness and community structure. W individuals were enriched for the mucin-degrading *Akkermansia muciniphila*. EA participants had increased levels of multiple *Bacteroidetes* species, fermentative pathways detected by metagenomics, and the short-chain fatty acid end products acetate, propionate, and isobutyrate. Differences in the gut microbiota between the EA and W groups could not be explained by reported dietary intake, were more pronounced for lean individuals, and were associated with current geographical location. Microbiome transplantations into germ-free mice confirmed that the differences in the gut microbiota of the EA and W individuals we analyzed are indeed independent of diet and that they differentially impact host body weight and adiposity in genetically identical mouse recipients.

Conclusions

The reported findings emphasize the utility of studying diverse ethnic groups within a defined geographical location and provide a starting point for dissecting the mechanisms contributing to the complex interactions between the gut microbiome and ethnicity-associated lifestyle, demographic, metabolic, and genetic factors.

Background

Culture-independent surveys have emphasized differences in gut microbial community structure between countries [1–3]; however, the factors that contribute to these differences are poorly understood. Diet is a common hypothesis for geographical variations in the gut microbiota [4, 5], based upon extensive data from intervention experiments in humans and mouse models [6–9]. However, diet is just one of the many factors that distinguishes human populations at the global scale, motivating the desire for a more holistic approach. Self-identified race/ethnicity (SIRE) provides a useful alternative, as it integrates the broader national or cultural tradition of a given social group. Multiple studies have reported associations between the gut microbiota and ethnicity in China [10], the Netherlands [11], Singapore [12], and the United States [13, 14]. In contrast, a recent study of Asian immigrants suggested that once an individual relocates to a new country, the microbiota rapidly assumes the structure of the country of residence [3]. Thus, the

degree to which microbiome signatures of ethnicity persist following immigration and their consequences for host pathophysiology remain an open question.

The links between ethnicity and metabolic disease are well-established. For example, EA subjects are more likely to develop health-related metabolic complications at lower body mass index (BMI) compared to their W counterparts [15, 16]. Moreover, Asian Americans have persistent ethnic differences in metabolic phenotypes following immigration [17], including a decoupling of BMI from total body fat percentage [18]. The mechanisms contributing to these ethnic differences in fat accrual remain unknown. Human genetic polymorphisms may play a role [19, 20]; however, putative alleles are often shared between members of different ethnic groups [21]. The gut microbiome might offer a possible explanation for differences in metabolic disease rates across ethnic groups, but there has been a relative scarcity of microbiome studies in this area.

These observations led us to hypothesize that ethnicity-associated differences in host metabolic phenotypes may be determined by corresponding differences in the gut microbiome. First, we sought to better understand the extent to which ethnicity is linked to the human gut microbiome in states of health and disease. We conducted a cross-sectional multi-omic analysis of the gut microbiome using paired 16S rRNA gene sequencing (16S-seq), metagenomics, and metabolomics from the Inflammation, Diabetes, Ethnicity, and Obesity (IDEO) cohort at the University of California, San Francisco. IDEO includes rich metabolic, dietary, and socioeconomic metadata [18], a restricted geographical distribution within the San Francisco Bay Area, and a balanced distribution of EA and W individuals that are both lean and obese (**Table S1**). We report marked differences in gut microbial richness, community structure, and metabolic end products between EA and W individuals in the IDEO cohort. We then used microbiome transplantations to assess the stability of ethnicity-associated differences in the gut microbiota in the context of genetically identical mice fed the same diet. We also explored the functional consequences of these differences on host metabolic phenotypes. Our results emphasize the importance of considering ethnicity in microbiome research and further complicate prior links between metabolic disease and the gut microbiome [22–24], which may be markedly different across diverse ethnic groups.

Methods

Human subjects

All study participants were part of the IDEO cohort, which has been previously described [18, 25]. Briefly, IDEO consists of 25–65 year-old men and women of multiple ethnicities and across a wide BMI range (18.5–52 kg/m²) living in the San Francisco Bay Area; exclusion factors include smoking, unstable weight within the last 3 months (> 3% weight gain or loss), a diagnosed inflammatory or infectious disease, liver failure, renal dysfunction, cancer, and reported alcohol consumption of > 20 grams per day. Using IDEO, we recruited both lean and obese W and EA individuals into this study based on World Health Organization cut-offs: W/EA BMI ≤ 24.9 kg/m² (lean); W BMI ≥ 30 kg/m² (obese); and EA BMI ≥ 27.5 kg/m² (obese) [17, 26, 27].

Each participant consented to take part in the study, which was approved by the University of California San Francisco (UCSF) Institutional Review Board. We utilized demographic, medical, dietary, and lifestyle metadata on each participant that were part of their initial recruitment into IDEO, as previously reported [18, 25, 28]. Participants with Type 2 Diabetes (T2D) were classified in accordance with American Diabetes Association Standards of Medical Care guidelines [29], defined by having glycated hemoglobin (HbA1c) $\geq 6.5\%$ or the combination of a prior diagnosis of T2D and the active use of an antidiabetic medication.

Anthropometric and body composition measurements

We leverage host phenotypic and demographic data from IDEO, which was the focus of two previous studies [18, 25]. For the convenience of the reader, we have reproduced our methods here. Height and weight were measured using a standard stadiometer and scale, and BMI (kg/m^2) was calculated from two averaged measurements. Waist and hip circumferences (to the nearest 0.5 cm) were measured using a plastic tape meter at the level of the umbilicus and of the greater trochanters, respectively, and waist-to-hip ratio (WHR) was calculated. Blood pressure was measured with a standard mercury sphygmomanometer on the left arm after at least 10 minute of rest. Mean values were determined from two independent measurements. Blood samples were collected after an overnight fast and analyzed for plasma glucose, insulin, serum total cholesterol, high density lipoprotein (HDL) cholesterol, and triglycerides. Low density lipoprotein (LDL) cholesterol was estimated according to the Friedewald formula [30]. Insulin resistance was estimated by the homeostatic model assessment of insulin resistance (HOMA-IR) index calculated from fasting glucose and insulin values [31]. Two obese subjects on insulin were included in the HOMA-IR analysis (1 EA, 1 W). Body composition of the subjects was estimated by Dual-Energy X-ray Absorptiometry (DEXA) using a Hologic Horizon/A scanner (3-minute whole-body scan, < 0.1 G milligray) per manufacturer protocol. A single technologist analyzed all DEXA measurements using Hologic Apex software (13.6.0.4:3) following the International Society for Clinical Densitometry guidelines. Visceral adipose tissue (VAT) was estimated from a 5 cm-wide region across the abdomen just above the iliac crest, coincident with the fourth lumbar vertebrae, to avoid interference from iliac crest bone pixels and matching the region commonly used to analyze VAT mass by CT scan [32–34].

Dietary assessment

IDEO participants completed two dietary questionnaires, as previously described [18, 25], allowing for the assessment of usual total fiber intake and fiber from specific sources, as well as macronutrient, phytochemical, vitamin, and mineral uptake. The first instrument was an Automated Self-Administered 24-hour Dietary Assessment (ASA24) [35, 36], which queries intake over a 24-hour period. The 24-hour recalls and supplement data were manually entered in the ASA24 Dietary Assessment Tool (v. 2016), an electronic data collection and dietary analysis program. ASA24 employs research-based strategies to enhance dietary recall using a respondent-driven approach allowing initial recall to be self-defined. The second instrument was the National Cancer Institute's Diet History Questionnaire III (DHQIII) [37, 38]. This questionnaire queries one's usual diet over the past month [38]. Completion of DHQIII added significantly

to participant survey fatigue, and completion rates were 42% for 1 recall, 79% for 2 recalls and 100% for 3 or 4 recalls after the first 5 months. Ultimately, our protocol was modified to request the completion of the ASA24 at three separate times, at appointments where there are computers and personnel assistance for online completion, as well as the DHQIII questionnaire.

DNA extraction

Human stool samples were homogenized with bead beating for 5 min (Mini-Beadbeater-96, BioSpec) using beads of mixed size and material (Lysing Matrix E 2 mL Tube, MP Biomedicals) in the digestion solution and lysis buffer of a Wizard SV 96 Genomic DNA kit (Promega). The samples were centrifuged for 10 min at 16,000 g and the supernatant was transferred to the binding plate. The DNA was then purified according to the manufacturer's instructions. Mouse fecal pellets were homogenized with bead beating for 5 min (Mini-Beadbeater-96, BioSpec) using the ZR BashingBead lysis matrix containing 0.1 and 0.5 mm beads (ZR-96 BashingBead Lysis Rack, Zymo Research) and the lysis solution provided in the ZymoBIOMICS 96 MagBead DNA Kit (Zymo Research). The samples were centrifuged for 5 min at 3,000 g and the supernatant was transferred to 1 mL deep-well plates. The DNA was then purified using the ZymoBIOMICS 96 MagBead DNA Kit (Zymo Research) according to the manufacturer's instructions.

16S rRNA gene sequencing and analysis

For human samples, 16S rRNA gene amplification was carried out using GoLay-barcoded 515F/806R primers [39] targeting the V4 region of the 16S rRNA gene according to the methods of the Earth Microbiome Project (earthmicrobiome.org) (**Table S2**). Briefly, 2 μ L of DNA was combined with 25 μ L of AmpliTaq Gold 360 Master Mix (Fisher Scientific), 5 μ L of primers (2 μ M each GoLay-barcoded 515/806R), and 18 μ L H₂O. Amplification was as follows: 10 min 95 °C, 30 $\times$ (30 s 95 °C, 30 s 50 °C, 30 s 72 °C), and 7 min 72 °C. Amplicons were quantified with PicoGreen (Quant-It dsDNA; Life Technologies) and pooled at equimolar concentrations. Aliquots of the pool were then column (MinElute PCR Purification Kit; Qiagen) and gel purified (QIAquick Gel Extraction Kit; Qiagen). Libraries were then quantified (KAPA Library Quantification Kit; Illumina) and sequenced with a 600 cycle MiSeq Reagent Kit (250 $\times$ 150; Illumina) with $\sim$ 15% PhiX spike-in. For mouse samples, 16S rRNA gene amplification was carried out as per reference protocol and primers [40]. In brief, the V4 region of the 16S rRNA gene was amplified with 515F/806R primers containing common adaptor sequences, and then the Illumina flow cell adaptors and dual indices were added in a secondary amplification step (see **Table S3** for index sequences). Amplicons were pooled and normalized using the SequalPrep Normalization Plate Kit (Invitrogen). Aliquots of the pool were then column (MinElute PCR Purification Kit, Qiagen) and gel purified (QIAquick Gel Extraction Kit, Qiagen). Libraries were then quantified and sequenced with a 600 cycle MiSeq Reagent Kit (250 $\times$ 250; Illumina) with $\sim$ 15% PhiX spike-in.

Demultiplexed sequencing reads were processed using QIIME2 v2020.2 (qiime2.org) with denoising by DADA2 [41]. Taxonomy was assigned using the DADA2 implementation of the RDP classifier [42] using the DADA2 formatted training sets for SILVA version 128 (benjjneb.github.io/dada2/assign.html). Amplicon sequence variants (ASVs) were filtered such that they were present in more than one sample

with at least a total of 10 reads across all samples. Alpha diversity metrics were calculated on subsampled reads using Vegan [43] and Picante [44] R packages. The PhILR Euclidean distance was calculated by first carrying out the phylogenetic isometric log ratio transformation (philr, PhILR [45]) followed by calculating the Euclidean distance (vegdist, Vegan [43]). Principal coordinates analysis was carried out using the pcoa function of APE [46]. ADONIS calculations were carried out (adonis, Vegan) with 999 replications on each distance metric. Centered log₂-ratio (CLR) normalized abundances were calculated using the Make.CLR function in MicrobeR package [47] with count zero multiplicative replacement (zCompositions; [48]). ALDEx2 [49] was used to analyze differential abundances of count data, using features that represented at least 0.05% of total sequencing reads. Corrections for multiple hypothesis testing to false discovery rate (FDR) using the Benjamin-Hochberg method [50] were performed where applicable. Analysis of distance matrices and alpha diversity mirror prior analyses developed in the Turnbaugh lab and were adapted to the current manuscript [9]. Calculations of associations between ASVs and ASA24 questionnaire data were completed by calculating a Spearman rank correlation and then adjusting the *p*-value for a Benjamini-Hochberg FDR using the cor_pmat function in the R package ggcorrplot.

Metagenomic sequencing and analysis

Whole-genome shotgun libraries were prepared using the Nextera XT DNA Library Prep Kit (Illumina). Paired ends of all libraries were sequenced on the NovaSeq 6000 platform in a single sequencing run (*n* = 45 subjects; see **Table S2** for relevant metadata and statistics). Illumina reads underwent quality trimming and adaptor removal using fastp [51] and host read removal using BMTagger v1.1.0 (<ftp.ncbi.nlm.nih.gov/pub/agarwala/bmtagger/>) in the metaWRAP pipeline (github.com/bxlab/metaWRAP). Metagenomic samples were taxonomically profiled using MetaPhlan2 v2.7.7 [52] and functionally profiled using HUMAnN2 v0.11.2 [53], both with default parameters. Principal coordinates analysis on MetaPhlan2 species-level abundances was carried out using Bray Curtis distances and the pcoa function of APE [46]. Tables of gene family abundances from HUMAnN2 were regrouped to KEGG orthologous groups using humann2_regroup_table. Functional pathways relating to short-chain fatty acid production were manually curated from the pathway outputs from HUMAnN2 and normalized by the estimated genome equivalents in each microbial community obtained from MicrobeCensus [54].

Quantification of bacterial load

Quantitative PCR (qPCR) was performed on DNA extracted from the human stool samples. DNA templates were diluted to 5 ng/μL before qPCR of total 16S rRNA gene copies was carried out in triplicate 10 μL reactions with 200 nM 340F/514R primers using a BioRad CFX384 thermocycler with SYBR Select for CFX Master Mix (Life Technologies) according to the manufacturer's instructions and an annealing temperature of 60°C. Absolute quantifications were determined against a standard curve of purified 8F/1542R amplified *Akkermansia muciniphila* DNA. Mean values of triplicate reactions were taken for further downstream analyses. Absolute bacterial abundance was derived by adjustments for dilutions

during DNA extraction and template normalization dividing by the total fecal mass used for DNA extraction in grams.

Nuclear magnetic resonance (NMR) metabolomics

NMR spectroscopy was performed at 298K on a Bruker Avance III 600 MHz spectrometer configured with a 5 mm inverse cryogenic probe (Bruker Biospin, Germany) as previously described [55]. 50 mg of human feces were extracted with 1 mL of phosphate buffer (K_2HPO_4/NaH_2PO_4 , 0.1 M, pH 7.4, 50% v/v D_2O) containing 0.005% sodium 3-(trimethylsilyl) [2,2,3,3- D_4] propionate (TSP- d_4) as a chemical shift reference (δ 0.00). Samples were freeze-thawed three times with liquid nitrogen and water bath for thorough extraction, then homogenized (6500 rpm, 1 cycle, 60 s) and centrifuged (11,180 g, 4 °C, 10 min). The supernatants were transferred to a new 2 mL tube. An additional 600 μ L of PBS was added to the pellets, followed by the same extraction procedure described above. Combined fecal extracts were centrifuged (11,180 g, 4 °C, 10 min), 600 μ L of the supernatant was transferred to a 5 mm NMR tube (Norell, Morganton, NC) for NMR spectroscopy analysis. A standard one-dimensional NOESY pulse sequence noesypr1d (recycle delay-90°-t1-90°-tm-90°-acquisition) is used with a 90 pulse length of approximately 10 s (-9.6 dbW) and 64 transients are recorded into 32 k data points with a spectral width of 9.6 KHz. NMR spectra were processed as previously described [55]. First, spectra quality was improved with Topspin 3.0 (Bruker Biospin, Germany) for phase and baseline correction and chemical shift calibration. AMIX software (version: 3.9.14, Bruker Biospin, Germany) was used for bucketing (bucket width 0.004 ppm), removal of interfering signal, and scaling (total intensity). Relative concentrations of identified metabolites were obtained by normalized peak area.

Targeted gas chromatography mass spectrometry (GC-MS) assays

Targeted analysis of short-chain fatty acids (SCFAs) and branched-chain amino acids (BCAAs) was performed with an Agilent 7890A gas chromatograph coupled with an Agilent 5975 mass spectrometer (Agilent Technologies Santa Clara, CA) using a propyl esterification method as previously described [55]. 50 mg of human fecal samples were pre-weighed, mixed with 1 mL of 0.005 M NaOH containing 10 μ g/mL caproic acid-6,6,6- d_3 (internal standard) and 1.0 mm diameter zirconia/silica beads (BioSpec, Bartlesville, OK). The mixture was thoroughly homogenized and centrifuged (13,200 g, 4 °C, 20 min). 500 μ L of supernatant was transferred to a 20 mL glass scintillation vial. 500 μ L of 1-propanol/pyridine (v/v = 3/2) solvent was added into the vial, followed by a slow adding of an aliquot of 100 μ L of esterification reagent propyl chloroformate. After a brief vortex of the mixture for 1 min, samples were derivatized at 60 °C for 1 hour. After derivatization, samples were extracted with hexane in a two-step procedure (300 μ L + 200 μ L) as described [56]. First, 300 μ L of hexane was added to the sample, briefly vortexed and centrifuged (2,000 g, 4 °C, 5 min), and 300 μ L of the upper layer was transferred to a glass autosampler vial. Second, an additional 200 μ L of hexane was added to the sample, vortexed, centrifuged, and the 200 μ L upper layer was transferred to the glass autosampler vial. A combination of 500 μ L of extracts were obtained for GC-MS analysis. A calibration curve of each SCFA and BCAA was generated with series

dilution of the standard for absolute quantitation of the biological concentration of SCFAs and BCAAs in human fecal samples.

Targeted bile acid quantitation by UPLC-MS/MS

Bile acid quantitation was performed with an ACQUITY ultra performance liquid chromatography (UPLC) system using a Ethylene Bridged Hybrid C8 column (1,7 μm , 100 mm x 2.1 mm) coupled with a Xevo TQ-S mass spectrometer equipped with an electrospray ionization (ESI) source operating in negative mode (All Waters, Milford, MA) as previously described [57]. Selected ion monitoring (SIM) for non-conjugated bile acids and multiple reaction monitoring (MRM) for conjugated bile acids was used. 50 mg of human fecal sample was pre-weighed, mixed with 1 mL of pre-cooled methanol containing 0.5 μM of stable-isotope-labeled bile acids (internal standards) and 1.0 mm diameter zirconia/silica beads (BioSpec, Bartlesville, OK), followed by thorough homogenization and centrifugation. Supernatant was transferred to an autosampler vial for analysis. 100 μL of serum was extracted by adding 200 μL pre-cooled methanol containing 0.5 μM deuterated bile acids as internal standards. Following centrifugation, the supernatant of the extract was transferred to an autosampler vial for quantitation. Calibration curves of individual bile acids were drafted with bile acid standards for quantitation of the biological abundance of bile acids.

Gnotobiotic mouse experiments

All mouse experiments were approved by the UCSF Institutional Animal Care and Use Committee and performed accordingly. Germ-free mice were maintained within the UCSF Gnotobiotic Core Facility and fed *ad libitum* autoclaved standard chow diet (Lab Diet 5021). Germ-free adult male C57BL/6J mice between 6–10 weeks of age were used for all the experiments described in this paper. 10 lean subjects in our IDEO cohort were selected as donors for the microbiota transplantation experiments, including 5 EA and 5 W donors. Donors matched for phenotypic data to the degree possible were selected (**Table S4**) for gnotobiotic experiments. Stool samples to be used for transplantation were resuspended in 10 volumes (by weight) of brain heart infusion media in an anaerobic Coy chamber. Each diluted sample was vortexed for 1 min and left to settle for 5 min, and a single 200 μL aliquot of the clarified supernatant was administered by oral gavage into each germ-free mouse recipient. In experiments LFPP1 and LFPP2, microbiome transplantations were performed for 2 donors per experiment (1 W, 1 EA) with gnotobiotic mice housed in sterile isolators (CBC flexible, softwall isolator) and maintained on *ad libitum* standard chow also known as low-fat, high-plant-polysaccharide (LFPP) diet. In LFPP1, 6 germ-free mice per colonization group received an aliquot of stool from a donor of either ethnicity and body composition (measured using EchoMRI) were recorded on the day of colonization and at 6 weeks post-transplantation (per group n = 6 recipient mice, 1 isolator, 2 cages). In LFPP2, we shortened the colonization time to 3 weeks and used two new donor samples. For the third experiment (HFHS experiment), mice were fed an irradiated high-fat, high-sugar diet (HFHS, TD.88137, Envigo) at weaning for four weeks prior to colonization and housed in pairs in Tecniplast IsoCages. We included the original 4 donor samples and included 6 new donors (per donor n = 2 recipient mice, 1 IsoCage). Body weight and body composition

were recorded on the day of colonization and at 3 weeks post-transplantation. Mice were maintained on the HFHS diet throughout the experiment. All samples were sequenced in a single pool (**Table S3**).

Glucose tolerance tests

Food was removed from mice 10 hr (LFPP1 experiment) or 4 hr (HFHS experiment) prior to assessment of glucose tolerance. Mice received i.p. injections of D-glucose (2 mg/kg), followed by repeated collection of blood by tail nick and determination of glucose levels by handheld glucometer (Abbot Diabetes Care) over a 2-hour period.

Geographic analyses

Map tiles and distance data was obtained using GGMap, OpenStreet Maps, and the lmap R packages. GGMap was employed using a Google Cloud API key and the final map tiles were obtained in July 2020 [58]. Spearman ranked correlation coefficients (ρ) were calculated as embedded in the R-package GGScatter and GGPlot2. 2018 US Census data for EA and W subjects was obtained (B02001 table for race, data.census.gov) for the ZIP codes available in our study and using the Leaflet package. The census data used is included as part of **Table S2** to aid in reproduction. Each census region is plotted as a percentage of W individuals over a denominator of W and EA subjects. The leaflet package utilized ZIP Code Tabulation Areas (ZCTAs) from the 2010 census. We extracted all ZCTAs starting with 9, and the resulting 29 ZIP codes that overlap with IDEO subjects were analyzed (**Table S2**). Two ZCTAs (95687 and 95401) were primarily W when comparing W and EA subjects. There were two W subjects recruited from these ZCTAs. These ZIP codes are cut off based on the zoom magnification for that figure and as a result ZCTAs for 27 individuals are plotted. Distance to a central point in SF was calculated. The point of reference was latitude = 37.7585102, longitude=-122.4539916. For the San Francisco Bay Bridge the point of reference was latitude = 37.7983, longitude= -122.3778.

Dietary questionnaire correlation analysis

DHQIII and ASA24 data were analyzed using a Euclidean distance matrix. These transformations were completed using the cluster package [59]. Subsequent analysis was completed using the vegan package [43]. Procrustes transformations were performed using 16S-seq data from human subjects, which was then subjected to a PhILR transformation. The resulting matrix was rotated against the distance matrix for ASA24 or DHQIII questionnaire data using the procrustes command in the vegan R package using 999 permutations. Mantel statistics were calculated utilizing the mantel command of the vegan package.

R packages used in this study

readxl [60], cluster [59], Rtsne [61], vegan [62], ape [63], ggpubr [64], leaflet [65], tigris [66], MicrobeR [47], OpenStreetMap [67], lmap [68], lmerTest [69], PhILR [45], qiime2R [70], ALDEx2 [49], gghighlight [71], Phyloseq [72], Janitor [73], table 1 [74], Picante [44], ggcorrplot [75].

Statistical analyses

Statistical analysis of the human data was performed using the `table1` package in R (STATCorp LLC, College Station, TX). Human data were presented as mean $\pm$ SD. Unpaired independent Student's *t* tests were used to compare differences between the two groups in the case of continuous data and in the case of categorical data the χ^2 test was utilized. These tests were adjusted for a Benjamini-Hochberg false discovery rate utilizing the command `p.adjust` in R, which is indicated as an adjusted *p*-value in the tables. In **Tables S8** and **S9** no values met an adjusted *p*-value cutoff of < 0.1 . In **Table S1**, *p*-values indicated by numbers were pooled together for adjustments and those represented by symbols were separately pooled together for adjustment. All microbiome-related analyses were carried out in R version 3.5.3 or 4.0.2. Where indicated, Wilcoxon rank-sum tests were calculated. A Benjamini-Hochberg false discovery rate of 0.1 was used as the cutoff for statistical significance was utilized in tables as indicated and an FDR of 0.2 was utilized for **Figures S5**. Statistical analysis of glucose tolerance tests was carried out using linear mixed effects models with the `lmerTest` R package and mouse as random effect. Graphical representation was carried out using `ggplot2`. Boxplots indicate the interquartile range (25th to 75th percentiles), with the center line indicating the median and whiskers representing 1.5x the interquartile range.

Results

Ethnicity was associated with marked differences in the human gut microbiota. Whereas there were no differences between ethnicities in estimated bacterial content per gram of stool (**Fig. S1A**), principal coordinates analysis of PhILR Euclidean distances from 16S-seq data (**Table S2**, *n* = 22 EA, 24 W subjects) revealed a subtle but significant separation between the gut microbiotas of EA and W subjects ($p = 0.037$, $R^2 = 0.037$, ADONIS; **Fig. 1A**). Statistical significance was robust to the distance metric used (**Table S5**). Bacterial diversity, evaluated using three distinct metrics from our 16S-seq data, were all significantly higher in W individuals, including Faith's phylogenetic diversity, ASV richness, and Shannon diversity (**Fig. 1B**). The *Firmicutes* and *Bacteroidetes* phyla were enriched in EA subjects, *Verrucomicrobia* were enriched in W subjects, and *Actinobacteria* and *Proteobacteria* were comparable (**Fig. 1C**). By contrast, analysis at the genus and ASV level did not reveal any differentially abundant groups between ethnicities, suggesting that the phylum-level trends we observed require the integration of more subtle shifts across multiple component members.

Next, we validated these results by metagenomic sequencing (**Table S2**, *n* = 21 EA, 24 W subjects). Consistent with our 16S-seq analysis, we detected a subtle but significant difference in the gut microbiomes between ethnicities based upon metagenomic species abundances ($p = 0.025$, $R^2 = 0.038$, ADONIS) and gene families ($p = 0.024$, $R^2 = 0.040$, ADONIS). Visualization of species within each phylum revealed marked variation in the magnitude and direction of change between ethnicities in our metagenomic (**Fig. 1D,E**) and 16S-seq data (**Fig. S1B,C**). Notably, W individuals had higher levels of *Akkermansia muciniphila* (**Fig. 1D,E** and **Fig. S1B,C**), which has been implicated in protection from obesity and its associated metabolic diseases [76].

Next, we used NMR-based stool metabolomics to gain insight into the potential functional consequences of ethnicity-associated differences in the human gut microbiome (**Table S2**, $n = 10$ subjects/ethnicity). Metabolite profiles were more strongly associated with ethnicity ($p = 0.008$, $R^2 = 0.128$, ADONIS; **Fig. 2A**) than community structure ($R^2 = 0.033-0.048$, ADONIS; **Table S5**) or gene abundance ($p = 0.024$, $R^2 = 0.040$, ADONIS). Feature annotations revealed elevated levels of the branched chain amino acid (BCAA) valine and the short-chain fatty acids (SCFAs) acetate and propionate in EA subjects (**Fig. 2B** and **Table S6**). In contrast, proline, formate, alanine, xanthine, and hypoxanthine were found at higher levels in W subjects (**Fig. 2B**). To assess the statistical significance and reproducibility of these trends, we used targeted GC-MS and UPLC-MS/MS to quantify a panel of BCAAs, SCFAs, and bile acids (**Table S7**). Confirming our NMR data, EA subjects had significantly higher levels of stool acetate (**Fig. 2C**) and propionate (**Fig. 2D**); however, we did not detect any significant differences in BCAAs or bile acids (**Fig. S2**). Isobutyrate (which was not detected by NMR) was also significantly higher in EA subjects (**Fig. 2E**). In agreement with these metabolite levels, a targeted re-analysis of our metagenomic data revealed a significant enrichment in two SCFA-related pathways: “pyruvate fermentation to butanoate” ($p = 0.023$, fold-difference = 2.216) and “superpathway of *Clostridium acetobutylicum* acidogenic fermentation” ($p = 0.023$, fold-difference = 2.182).

Consistent with prior studies [22, 23, 77], we found that gut bacterial richness in W individuals was significantly associated with both BMI (**Fig. 3A**) and body fat percentage (**Fig. 3B**). Remarkably, these associations were undetectable in EA subjects (**Figs. 3A,B**) even when other metrics of bacterial diversity were used (**Fig. S3**). Re-analysis of our data separating lean and obese individuals revealed that the previously observed differences between ethnic groups were driven by lean individuals. Lean W subjects had significantly higher bacterial diversity (**Fig. 3C**), in addition to greater differences in both gut microbial community structure ($p = 0.001$, $R^2 = 0.096$, ADONIS; **Fig. 3D**) and metabolite profiles ($p = 0.006$, $R^2 = 0.293$, ADONIS; **Fig. 3E**), than did corresponding EA individuals. By contrast, obese W vs. EA individuals were not different across any of these metrics (**Figs. 3C-E**). Lean EA individuals were significantly enriched for the *Actinobacteria* and *Firmicutes* phyla with a trend towards increased *Bacteroidetes* (**Fig. 3F**). At the genus level, lean EA subjects had higher levels of *Bacteroides*, *Blautia*, and an unclassified *Lachnospiraceae* taxon (**Figs. 3G,H**). In contrast, the *Verrucomicrobia* phylum (which contains *A. muciniphila*) was consistently enriched in both lean and obese W subjects relative to EA individuals (**Fig. 3F**).

Next, we sought to understand the potential drivers of differences in the gut microbiome between ethnic groups in lean individuals within the IDEO cohort, focusing on birth location, time spent in the USA, dietary intake, and host metabolic phenotypes. Although everyone in the cohort was recruited from the San Francisco Bay Area, birth location varied widely (**Fig. S4**). There was no significant difference in the proportion of subjects born in the USA between ethnicities (75% W, 54.5% EA; $p = 0.15$, Pearson's χ^2 test). There was also no significant difference in the geographical distance between birth location and San Francisco [W median 2,318 (2.2-6,906) miles; EA median 1,986 (2.2-6,906) miles; $p = 0.69$, Wilcoxon rank-sum test) or the amount of time spent in the San Francisco Bay Area at the time of sampling [W median

270 (8.00-741) months; EA median 282.5 (8.50–777) years; $p = 0.42$, Wilcoxon rank-sum test). While obese subjects were markedly distinct from lean individuals of both ethnicities with regard to measured metabolic and laboratory parameters (**Table S1**), there were no statistically significant differences between ethnic groups after separating lean and obese individuals (**Table S1**).

Surprisingly, we did not detect any significant differences in either short- (**Table S8**) or long-term (**Table S9**) dietary intake between ethnicities. Consistent with this, Procrustes analysis did not reveal any significant associations between dietary intake and gut microbial community structure: procrustes $p = 0.452$ (DHQIII) and $p = 0.445$ (ASA24) relative to PhILR transformed 16S-seq ASV data. The Spearman Mantel statistic was also non-significant [$r = 0.09511$, $p = 0.094$ (DHQIII) and $r = 0.02953$, $p = 0.313$ (ASA24)], relative to PhILR transformed 16S-seq ASV data. Despite the lack of a strong overall shift in the gut microbiota, we were able to identify 11 ASVs associated with dietary intake in lean W individuals (**Fig. S5**). In contrast, there were no significant ASV-level associations in lean EA subjects.

Given the marked variation in the gut microbiome at the continental scale [1–3], we hypothesized that the observed differences in lean EA and W individuals may be influenced by a participant's current address at the time of sampling. Consistent with this hypothesis, we found clear trends in ethnic group composition across ZIP codes in the IDEO cohort (**Figs. 4A,B**) that were mirrored by the 2018 US census data (Pearson $r = 0.52$, $p = 0.026$ for neighborhoods with greater than 50% white subjects; **Fig. 4D**). Obese individuals from both ethnicities and lean W subjects tended to live closer to the center of San Francisco relative to lean EA subjects (**Fig. 4C**). Distance between current ZIP code and the center of San Francisco was associated with both gut microbial diversity (**Fig. 4E**) and community structure (**Fig. 4F**). These analyses were robust to the central point used, as shown using the Bay Bridge as the central reference point (**Fig. S6**).

Taken together, our results support the hypothesis that there are stable ethnicity-associated signatures within the gut microbiota of lean EA vs. W individuals that are independent of diet. To experimentally test this hypothesis, we transplanted the gut microbiota of a representative lean W and lean EA individual into germ-free C57BL/6J mice (**Fig. 5A** and **Table S4**). Despite maintaining the genetically identical recipient mice on the same autoclaved low-fat, high plant-polysaccharide (LFPP) diet, we detected significant differences in gut microbial community structure (**Fig. 5B**), bacterial richness (**Fig. 5C**), and taxonomic abundance (**Figs. 5D-E**) between the two ethnicity-specific recipient groups, reflecting the differences we saw when initially analyzing the stool samples of the human participants themselves. A replication experiment using an independent pair of donors revealed similar trends, though they did not reach statistical significance (**Fig. S7**). To assess the reproducibility of these findings across multiple donors and in the context of a distinctive dietary pressure, we next fed 20 germ-free mice a high-fat, high-sugar (HFHS) diet for 4 weeks prior to colonization with microbiota from a W vs. EA donor, and then maintained the mice on this diet following colonization (per group $n = 10$ mice, 5 donors; per donor $n = 2$ mice, 1 cage; **Fig. 5F**). Remarkably, this experiment replicated our original findings on the LFPP diet, including altered gut microbial community structure (**Fig. 5G**), increased richness in mice receiving W donor microbiota (**Fig. 5H**), and higher levels of *Bacteroides* in mice receiving EA donor gut microbiota (**Figs. 5I-J**).

Moreover, mice transplanted with gut microbiomes of EA and W individuals displayed significant differences in body composition. Mice that received W donor microbiota and were fed the LFPP diet gained more body weight (**Fig. 6A**) and increased their adiposity (**Fig. 6B**), in conjunction with a reduction in lean mass (**Fig. 6C**), relative to mice that received the EA donor microbiota. These overall trends were mirrored in the human microbiota recipient mice that were fed the HFHS diet (**Figs. 6E-G**), though they did not reach statistical significance due to marked variations between donors independent of ethnicity (**Fig. S8**). There were no significant differences in glucose tolerance between mice receiving stool transplants from donors of different ethnicity in either experiment (**Figs. 6D,H**).

Discussion

Our findings suggest that despite the potential for immigration to erase some of the geographical structure in the gut microbiome [3], there remain stable long-lasting microbial signatures of ethnicity among W and EA residents of the San Francisco Bay Area. The mechanisms responsible remain to be elucidated. In lean individuals within the IDEO cohort, these differences appear to be independent of immigration status, host phenotype, or dietary intake. Our experiments in inbred germ-free mice support the stability of ethnicity-associated differences in the gut microbiota on both the LFPP and HFHS diets, while also demonstrating that variations in host genetics are not necessary to maintain these signatures, at least over short timescales. Our data also supports a potential role for geographic location of residence in reinforcing differences in the gut microbiota between ethnic groups. The specific reasons why current location matters remains unclear. It may be reflective of subtle differences in dietary intake (*e.g.*, ethnic foods, food source, or phytochemical content) that are hard to capture using the validated nutritional surveys [78]. Alternative hypotheses include biogeographical patterns in microbial dispersion [79] or a role for socioeconomic factors, which are correlated with neighborhood [80].

Surprisingly, our findings demonstrate that ethnicity-associated differences in the gut microbiota are stronger in lean individuals. Obese individuals did not exhibit as clear a difference in the gut microbiota between ethnic groups, either suggesting that established obesity can overwrite long-lasting microbial signatures or alternatively that there is a shared ethnicity-independent microbiome type that predisposes individuals to obesity. Studies in other disease areas (*e.g.*, inflammatory bowel disease and cancer) with similar multi-ethnic cohorts are essential to test the generalizability of these findings and to generate hypotheses as to their mechanistic underpinnings.

Our results in humans and mouse models support the broad potential for downstream consequences of ethnicity-associated differences in the gut microbiome for metabolic syndrome and potentially other disease areas. However, the causal relationships and how they can be understood in the context of the broader differences in host phenotype between ethnicities require further study. While these data are consistent with our general hypothesis that ethnicity-associated differences in the gut microbiome are a source of differences in host metabolic disease, we were surprised by both the nature of the microbiome shifts and their directionality. Based upon observations in the IDEO [18] and other cohorts [15, 16], we anticipated that the gut microbiomes of lean EA individuals would promote obesity or other features of

metabolic syndrome. In humans, we did find multiple signals that have been previously linked to obesity and its associated metabolic diseases in EA individuals, including increased Firmicutes [9, 81], decreased *A. muciniphila* [76, 82], decreased diversity [23], and increased acetate [83, 84]. Yet EA subjects also had higher levels of *Bacteroidetes* and *Bacteroides*, which have been linked to improved metabolic health [85]. More importantly, our microbiome transplantations demonstrated that the recipients of the lean EA gut microbiome had less body fat despite consuming the same LFPP and HFHS diets. These seemingly contradictory findings may suggest that the recipient mice lost some of the microbial features of ethnicity relevant to host metabolic disease or alternatively that the microbiome acts in a beneficial manner to counteract other ethnicity-associated factors driving disease.

EA subjects also had elevated levels of the short chain fatty acids propionate and isobutyrate. The consequences of elevated intestinal propionate levels are unclear given the seemingly conflicting evidence in the literature that propionate may either exacerbate [86] or protect [87] from aspects of metabolic syndrome. Clinical data suggests that circulating propionate may be more relevant for disease than fecal levels [88], emphasizing the importance of considering both the specific microbial metabolites produced, their intestinal absorption, and their distribution throughout the body. Isobutyrate is even less well-characterized, with prior links to dietary intake [89] but no association with obesity [90].

There are multiple limitations of this study. Due to the investment of resources into ensuring a high level of phenotypic information on each cohort member, and due to its restricted geographical catchment area, the IDEO cohort was relatively small at the time of this analysis ($n = 46$ individuals). This study only focused on two of the major ethnicities in the San Francisco Bay Area; as IDEO continues to expand and diversify its membership, we hope to study a sufficient number of participants from other ethnic groups in the future. Stool samples were collected at a single time point and analyzed in a cross-sectional manner. While we used validated tools from the field of nutrition to monitor dietary intake, we cannot fully exclude subtle dietary differences between ethnicities [91], which would be possible with a controlled feeding study [81]. In our animal studies, we focused on metabolically healthy wild-type germ-free mice. Follow-up studies in the appropriate disease models coupled to controlled experimentation with individual strains or more complex synthetic communities are necessary to elucidate the mechanisms responsible for ethnicity-associated changes in host physiology and their relevance to disease.

Conclusions

Our results support the utility of considering ethnicity as a covariate in microbiome studies, due to the ability to detect signals that are difficult to capture by more specific metadata such as individual dietary intake values. On the other hand, these findings raise the importance of dissecting the sociological and biological components of ethnicity with the goal of identifying factors that shape the gut microbiota, either alone or in combination. This emerging area of microbiome research is just one component in the broader efforts to explore the boundaries and mechanistic underpinning of ethnicity with respect to multiple ethnic groups. The IDEO cohort provides a valuable research tool to conduct prospective longitudinal and intervention studies examining diabetes including Asian American participants. More

broadly, IDEO provides a framework to approach other disease states where self-identified race or ethnicity are thought to contribute to health outcomes related to the microbiome. By understanding the biologic features that drive differences between ethnic groups, we may be able to achieve similar health outcomes and to support more precise therapies informed by a broader appreciation of both microbial and human diversity.

Abbreviations

16S-seq: 16S rRNA gene sequencing

ASA24: Automated Self Administered 24-Hour Dietary Assessment Tool

ASV: Amplicon Sequence Variant

BCAA: branched chain amino acid

BMI: body mass index

DHQIII: Diet History Questionnaire III

EA: East Asian

HFHS: high-fat, high-sugar

LFPP: low-fat, plant-polysaccharide rich

SCFA: short-chain fatty acid

W: White

ZCTA: ZIP Code Tabulation Area

Declarations

Ethics approval and consent to participate

Human stool samples were collected as part of a multi-ethnic clinical cohort study termed Inflammation, Diabetes, Ethnicity and Obesity (ClinicalTrials.gov identifier NCT03022682), consisting of 25- to 65-year-old men and women residing in Northern California and recruited from medical and surgical clinics at UCSF and the Zuckerberg San Francisco General Hospital, or through local public advertisements. The host phenotypic data from this cohort have been described in detail [\[18,25\]](#). Informed consent was provided for all subjects participating in the study, which was approved by the UCSF Institutional Review Board. Protocols for all experiments involving mice were approved by the University of California, San Francisco Institutional Animal Care and Use Committee, and performed accordingly.

Consent for publication

Not applicable.

Availability of data and materials

All 16S-seq and metagenomic sequencing data generated in the preparation of this manuscript have been deposited in NCBI's Sequence Read Archive under accession number PRJNA665061. Metabolomics results and metadata are available within this manuscript (**Tables S2, S3, S5, and S6**). Code for our manuscript will be uploaded to GitHub (<https://github.com/turnbaughlab>).

Competing interests

P.J.T. is on the scientific advisory board for Kaleido, Pendulum, Seres, and SNIPRbiome; there is no direct overlap between the current study and these consulting duties. All other authors declare that they have no competing interests.

Funding

This work was supported by the National Institutes of Health [R01HL122593, R21CA227232, R01AR074500, R01DK114034 (PJT); 1R01DK11230401, R01DK11230403S1, (SKK)]. PJT is a Chan Zuckerberg Biohub investigator and was a Nadia's Gift Foundation Innovator, supported in part by the Damon Runyon Cancer Research Foundation (DRR-42-16). DLA is supported by the American Diabetes Association (1-18-PMF-003). VU is supported by the National Institutes of Health [T32HL007185](#). ADP is supported in part by the Pennsylvania Department of Health using Tobacco CURE funds, and the USDA National Institute of Food and Federal Appropriations under Project PEN04607 and Accession number 1009993. None of these funding bodies played a role in the design of the study, the collection, analysis, and interpretation of data, or in writing the manuscript.

Authors' contributions

PJT and SKK designed the study. DLA and SKK recruited the IDEO cohort. QYA and DLA performed all gnotobiotic mouse experiments. QYA and EB performed DNA extractions and library preparations for sequencing. QYA and VU conducted all bioinformatic analyses including analysis of microbiome, geographic, and dietary datasets. JEB wrote the bioinformatic pipelines for processing 16S-seq and metagenomic data and assisted with data analysis and interpretation. HLL, GW, and EB helped with the annotation and analysis of subjects' metadata and dietary questionnaires. JC and ADP conducted and analyzed metabolomics data. PJT, QYA, and VU wrote the manuscript and prepared the figures with input from all co-authors.

Acknowledgements

We thank Jessie Turnbaugh and the other UCSF Gnotobiotics Core Facility staff and members of the Koliwad lab for help with the gnotobiotic mouse experiments. We thank the Penn State Metabolomics

Facility. We also thank the CZ Biohub Sequencing Platform for sequencing support, as well as all the subjects who participated in this study.

References

1. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature*. Nature Publishing Group; 2012;486:222–7.
2. Hehemann J-H, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature*. Nature Publishing Group; 2010;464:908–12.
3. Vangay P, Johnson AJ, Ward TL, Al-Ghalith GA, Shields-Cutler RR, Hillmann BM, et al. US Immigration Westernizes the Human Gut Microbiome. *Cell*. Nature Publishing Group; 2018;175:962–72.e10.
4. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A*. 2010;107:14691–6.
5. Devoto AE, Santini JM, Olm MR, Anantharaman K, Munk P, Tung J, et al. Megaphages infect *Prevotella* and variants are widespread in gut microbiomes. *Nat Microbiol*. 2019;4:693–700.
6. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. 2014;505:559–63.
7. Carmody RN, Gerber GK, Luevano JM Jr, Gatti DM, Somes L, Svenson KL, et al. Diet dominates host genotype in shaping the murine gut microbiota. *Cell Host Microbe*. 2015;17:72–84.
8. Gehrig JL, Venkatesh S, Chang H-W, Hibberd MC, Kung VL, Cheng J, et al. Effects of microbiota-directed foods in gnotobiotic animals and undernourished children. *Science*. 2019;365.
9. Bisanz JE, Upadhyay V, Turnbaugh JA, Ly K, Turnbaugh PJ. Meta-Analysis Reveals Reproducible Gut Microbiome Alterations in Response to a High-Fat Diet. *Cell Host Microbe*. 2019;26:265–72.e4.
10. Khine WWT, Zhang Y, Goie GJY, Wong MS, Liong M, Lee YY, et al. Gut microbiome of pre-adolescent children of two ethnicities residing in three distant cities. *Sci Rep*. 2019;9:7831.
11. Deschasaux M, Bouter KE, Prodan A, Levin E, Groen AK, Herrema H, et al. Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. *Nat Med*. 2018;24:1526–31.
12. Xu J, Lawley B, Wong G, Otal A, Chen L, Ying TJ, et al. Ethnic diversity in infant gut microbiota is apparent before the introduction of complementary diets. *Gut Microbes*. 2020;11:1362–73.
13. Brooks AW, Priya S, Blekhman R, Bordenstein SR. Gut microbiota diversity across ethnicities in the United States. *PLoS Biol*. 2018;16:e2006842.
14. Sordillo JE, Zhou Y, McGeachie MJ, Ziniti J, Lange N, Laranjo N, et al. Factors influencing the infant gut microbiome at age 3-6 months: Findings from the ethnically diverse Vitamin D Antenatal Asthma Reduction Trial (VDAART). *J Allergy Clin Immunol*. Elsevier; 2017;139:482–91.e14.

15. Zheng W, McLerran DF, Rolland B, Zhang X, Inoue M, Matsuo K, et al. Association between body-mass index and risk of death in more than 1 million Asians. *N Engl J Med*. Mass Medical Soc; 2011;364:719–29.
16. Gu D, He J, Duan X, Reynolds K, Wu X, Chen J, et al. Body weight and mortality among men and women in China. *JAMA*. jamanetwork.com; 2006;295:776–83.
17. Jih J, Mukherjea A, Vittinghoff E, Nguyen TT, Tsoh JY, Fukuoka Y, et al. Using appropriate body mass index cut points for overweight and obesity among Asian Americans. *Prev Med* . Elsevier; 2014;65:1–6.
18. Alba DL, Farooq JA, Lin MYC, Schafer AL, Shepherd J, Koliwad SK. Subcutaneous Fat Fibrosis Links Obesity to Insulin Resistance in Chinese Americans. *J Clin Endocrinol Metab*. academic.oup.com; 2018;103:3194–204.
19. Xiang K, Wang Y, Zheng T, Jia W, Li J, Chen L, et al. Genome-wide search for type 2 diabetes/impaired glucose homeostasis susceptibility genes in the Chinese: significant linkage to chromosome 6q21-q23 and chromosome 1q21-q24. *Diabetes*. Am Diabetes Assoc; 2004;53:228–34.
20. Wen J, Rönn T, Olsson A, Yang Z, Lu B, Du Y, et al. Investigation of type 2 diabetes risk alleles support CDKN2A/B, CDKAL1, and TCF7L2 as susceptibility genes in a Han Chinese cohort. *PLoS One*. journals.plos.org; 2010;5:e9153.
21. Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, et al. Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences*. National Acad Sciences; 2011;108:11983–8.
22. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Human gut microbes associated with obesity. *Nature*. 2006;444:1022–3.
23. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. *Nature*. 2009;457:480–4.
24. Wu H, Tremaroli V, Schmidt C, Lundqvist A, Olsson LM, Krämer M, et al. The Gut Microbiota in Prediabetes and Diabetes: A Population-Based Cross-Sectional Study. *Cell Metab*. 2020;32:379–90.
25. Oguri Y, Shinoda K, Kim H, Alba DL, Bolus WR, Wang Q, et al. CD81 Controls Beige Fat Progenitor Cell Growth and Energy Balance via FAK Signaling. *Cell*. 2020;182:563–77.e20.
26. Hsu WC, Araneta MRG, Kanaya AM, Chiang JL, Fujimoto W. BMI cut points to identify at-risk Asian Americans for type 2 diabetes screening. *Diabetes Care*. Am Diabetes Assoc; 2015;38:150–8.
27. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet*. 2004;363:157–63.
28. McClung HL, Ptomey LT, Shook RP, Aggarwal A, Gorczyca AM, Sazonov ES, et al. Dietary Intake and Physical Activity Assessment: Current Tools, Techniques, and Technologies for Use in Adult Populations. *Am J Prev Med*. Elsevier; 2018;55:e93–104.
29. American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2019. *Diabetes Care*. American Diabetes Association; 2019;42:S13–28.

30. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem. academic.oup.com*; 1972;18:499–502.
31. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia. Springer*; 1985;28:412–9.
32. Kaul S, Rothney MP, Peters DM, Wacker WK, Davis CE, Shapiro MD, et al. Dual-Energy X-Ray Absorptiometry for Quantification of Visceral Fat . *Obesity*. 2012. p. 1313–8.
33. Bredella MA, Gill CM, Keating LK, Torriani M, Anderson EJ, Punyanitya M, et al. Assessment of abdominal fat compartments using DXA in premenopausal women from anorexia nervosa to morbid obesity. *Obesity* . 2013;21:2458–64.
34. Neeland IJ, Grundy SM, Li X, Adams-Huet B, Vega GL. Comparison of visceral fat mass measurement by dual-X-ray absorptiometry and magnetic resonance imaging in a multiethnic cohort: the Dallas Heart Study. *Nutr Diabetes. nature.com*; 2016;6:e221.
35. Timon CM, van den Barg R, Blain RJ, Kehoe L, Evans K, Walton J, et al. A review of the design and validation of web- and computer-based 24-h dietary recall tools . *Nutrition Research Reviews*. 2016. p. 268–80.
36. Park Y, Dodd KW, Kipnis V, Thompson FE, Potischman N, Schoeller DA, et al. Comparison of self-reported dietary intakes from the Automated Self-Administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers . *The American Journal of Clinical Nutrition*. 2018. p. 80–93.
37. Millen AE, Midthune D, Thompson FE, Kipnis V, Subar AF. The National Cancer Institute diet history questionnaire: validation of pyramid food servings. *Am J Epidemiol*. 2006;163:279–88.
38. Diet History Questionnaire III (DHQ III). Available from: <https://epi.grants.cancer.gov/dhq3/>
39. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*. 2012;6:1621–4.
40. Gohl DM, Vangay P, Garbe J, MacLean A, Hauge A, Becker A, et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat Biotechnol*. 2016;34:942–9.
41. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13:581–3.
42. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*. 2007;73:5261–7.
43. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci*. 2003;14:927–30.
44. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, et al. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*. 2010;26:1463–4.

45. Silverman JD, Washburne AD, Mukherjee S, David LA. A phylogenetic transform enhances analysis of compositional microbiota data. *Elife* . elifesciences.org; 2017;6.
46. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. 2004;20:289–90.
47. Bisanz J. MicrobeR v 0.3.2: Handy functions for microbiome analysis in R. *GitHub* . 2017.
48. Martín-Fernández J-A, Hron K, Templ M, Filzmoser P, Palarea-Albaladejo J. Bayesian-multiplicative treatment of count zeros in compositional data sets. *Stat Modelling*. SAGE Publications India; 2015;15:134–58.
49. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-like differential gene expression analysis of single-organism and meta-RNA-seq. *PLoS One*. 2013;8:e67019.
50. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. Royal Statistical Society, Wiley; 1995;57:289–300.
51. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:i884–90.
52. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods*. 2012;9:811–4.
53. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods*. 2018;15:962–8.
54. Nayfach S, Pollard KS. Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol*. 2015;16:51.
55. Cai J, Zhang J, Tian Y, Zhang L, Hatzakis E, Krausz KW, et al. Orthogonal comparison of GC-MS and ¹H NMR spectroscopy for short chain fatty acid quantitation. *Anal Chem*. 2017;89:7900–6.
56. Zheng X, Qiu Y, Zhong W, Baxter S, Su M, Li Q, et al. A targeted metabolomic protocol for short-chain fatty acids and branched-chain amino acids. *Metabolomics*. 2013;9:818–27.
57. Sarafian MH, Lewis MR, Pechlivanis A, Ralphs S, McPhail MJW, Patel VC, et al. Bile acid profiling and quantification in biofluids using ultra-performance liquid chromatography tandem mass spectrometry. *Anal Chem*. 2015;87:9662–70.
58. Kahle D, Wickham H. ggmap: Spatial Visualization with ggplot2. *R J*. researchgate.net; 2013;5:144–61.
59. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K, Others. Cluster: cluster analysis basics and extensions. R package version. 2012;1:56.
60. Wickham H, Bryan J. readxl: Read Excel Files. R package version 1.0. 0. URL <https://CRAN.R-project.org/package=readxl>. 2017.
61. Krijthe JH. Rtsne: T-distributed stochastic neighbor embedding using Barnes-Hut implementation. R package version 0.13, URL <https://github.com/jkrijthe/Rtsne>. 2015.

62. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara RB, et al. Community ecology package. R package version. sortie-nd.org; 2013;2–0.
63. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics . academic.oup.com; 2019.
64. Kassambara A. ggpubr:“ggplot2” based publication ready plots. R package version 0 1. 2018;7.
65. Cheng J, Karambelkar B, Xie Y. Leaflet: Create interactive web maps with the javascript'leaflet'library. R package version. 2018;2.
66. Walker K. Tigris: Load census TIGER/Line Shapefiles. R package version 0.7. 2018.
67. Fellows I, Stotz JP. OpenStreetMap: Access to open street map raster images. R Package Version, 0.3. 3. 2016.
68. Wallace JR. Imap: Interactive mapping. R package version 1.32. R Found Stat Comput; 2012.
69. Kuznetsova A, Brockhoff PB, Christensen RHB, Others. lmerTest package: tests in linear mixed effects models. J Stat Softw. core.ac.uk; 2017;82:1–26.
70. Bisanz JE. qiime2R: Importing QIIME2 artifacts and associated data into R sessions. Version 0 99. 2018;13.
71. Yutani H. gghighlight: Highlight Lines and Points in “ggplot2.” Manual available online at <http://CRAN.R-project.org/package=gghighlight>. 2018.
72. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. journals.plos.org; 2013;8:e61217.
73. Firke S. Janitor: Simple tools for examining and cleaning dirty data. R package version. 2018;1.
74. Rich B. table1: Tables of Descriptive Statistics in HTML. R package version 12 . 2020.
75. Kassambara A, Kassambara MA. Package “ggcorrplot.” R package version 0 1 . mran.microsoft.com; 2019;3.
76. Plovier H, Everard A, Druart C, Depommier C, Van Hul M, Geurts L, et al. A purified membrane protein from Akkermansia muciniphila or the pasteurized bacterium improves metabolism in obese and diabetic mice. Nat Med. 2017;23:107–13.
77. Jumpertz R, Le DS, Turnbaugh PJ, Trinidad C, Bogardus C, Gordon JI, et al. Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. Am J Clin Nutr. 2011;94:58–65.
78. Garduño-Díaz SD, Husain W, Ashkanani F, Khokhar S. Meeting challenges related to the dietary assessment of ethnic minority populations. J Hum Nutr Diet. Wiley Online Library; 2014;27:358–66.
79. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, et al. Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol. nature.com; 2006;4:102–12.
80. Kakar V, Voelz J, Wu J, Franco J. The Visible Host: Does race guide Airbnb rental rates in San Francisco? J Hous Econ. Elsevier; 2018;40:25–40.
81. Basolo A, Hohenadel M, Ang QY, Piaggi P, Heinitz S, Walter M, et al. Effects of underfeeding and oral vancomycin on gut microbiome and nutrient absorption in humans. Nat Med. 2020;26:589–98.

82. Depommier C, Everard A, Druart C, Plovier H, Van Hul M, Vieira-Silva S, et al. Supplementation with *Akkermansia muciniphila* in overweight and obese human volunteers: a proof-of-concept exploratory study. *Nat Med*. 2019;25:1096–103.
83. Perry RJ, Peng L, Barry NA, Cline GW, Zhang D, Cardone RL, et al. Acetate mediates a microbiome–brain– β -cell axis to promote metabolic syndrome. *Nature*. Nature Publishing Group; 2016;534:213–7.
84. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*. 2006;444:1027–31.
85. Johnson EL, Heaver SL, Walters WA, Ley RE. Microbiome and metabolic disease: revisiting the bacterial phylum Bacteroidetes. *J Mol Med* . Springer; 2017;95:1–8.
86. Tirosh A, Calay ES, Tuncman G, Claiborn KC, Inouye KE, Eguchi K, et al. The short-chain fatty acid propionate increases glucagon and FABP4 production, impairing insulin action in mice and humans. *Sci Transl Med* . stm.sciencemag.org; 2019;11.
87. Lu Y, Fan C, Li P, Lu Y, Chang X, Qi K. Short Chain Fatty Acids Prevent High-fat-diet-induced Obesity in Mice by Regulating G Protein-coupled Receptors and Gut Microbiota. *Sci Rep*. nature.com; 2016;6:37589.
88. Müller M, Hernández MAG, Goossens GH, Reijnders D, Holst JJ, Jocken JWE, et al. Circulating but not faecal short-chain fatty acids are related to insulin sensitivity, lipolysis and GLP-1 concentrations in humans. *Sci Rep*. nature.com; 2019;9:12515.
89. Berding K, Donovan SM. Diet Can Impact Microbiota Composition in Children With Autism Spectrum Disorder. *Front Neurosci*. frontiersin.org; 2018;12:515.
90. Kim KN, Yao Y, Ju SY. Short Chain Fatty Acids and Fecal Microbiota Abundance in Humans with Obesity: A Systematic Review and Meta-Analysis. *Nutrients* . mdpi.com; 2019;11.
91. Johnson AJ, Vangay P, Al-Ghalith GA, Hillmann BM, Ward TL, Shields-Cutler RR, et al. Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans. *Cell Host Microbe*. 2019;25:789–802.e5.

Figures

Figure 1

The gut microbiota is distinct between East Asian and White subjects living in the Bay Area. (A-C) 16S-seq data for subjects in the IDEO cohort. Each point represents a single individual's gut microbiota ($n=22$ EA, $n=24$ W). (A) Principal coordinate analysis of PhILR Euclidean distances reveals significant separation between ethnic groups ($p=0.037$, $R^2=0.037$, ADONIS). Additional distance metrics are shown in Table S5. (B) Microbial diversity metrics are significantly different between EA and W subjects. p -values determined using Wilcoxon rank-sum tests. (C) Three of the five most abundant bacterial phyla are significantly different between EA and W individuals (p -values determined using Wilcoxon rank-sum test). (D,E) Metagenomic data ($n=21$ EA, $n=24$ W). (D) Each point represents the average relative abundance for a

given species within each ethnic group, connected with a line that is colored by the ethnic group with higher mean abundance of each species: EA (yellow) and W (purple). (E) Pie charts indicate the proportion of species within each phylum that have greater mean abundance in EA or W subjects.

Figure 2

Metabolomics and targeted mass spectrometry highlight significant differences in bacterial fermentation end-products between ethnicities. (A) Global profiling of the stool metabolome by proton nuclear magnetic resonance (^1H NMR) revealed a significant separation in metabolomic profiles between EA and W individuals ($p=0.008$, $R^2=0.128$, ADONIS). (B) Representative stool metabolites contributing to the separation of stool metabolomic profiles between EA and W individuals ($p<0.05$, Wilcoxon rank-sum test). (C-E) Gas chromatography-mass spectrometry analysis of short-chain fatty acids (SCFAs) revealed significantly higher concentrations of acetate (C), propionate (D) and isobutyrate (E) in the stool samples of EA compared to W individuals. p -values determined using Wilcoxon rank-sum tests. (A-E) $n=10$ EA and $n=10$ W individuals.

Figure 3

Ethnicity-associated differences in gut microbial community structure are restricted to lean individuals. (A,B) Faith's phylogenetic diversity is negatively correlated with (A) BMI and (B) percent body fat in W but not EA individuals (p-values and Spearman rank correlation coefficients are shown for each graph). (C) Microbial diversity metrics were significantly different between lean EA and W individuals, but not for obese subjects of the two ethnicities. p-values determined using Wilcoxon rank-sum tests. (D) Principal

coordinate analysis of PhILR Euclidean distances reveals significant separation between the gut microbiotas of EA and W lean individuals ($p=0.001$, $R^2=0.096$, ADONIS), with no separation in obese subjects ($p=0.7$, $R^2=0.036$, ADONIS). (E) Global profiling of the stool metabolome by proton nuclear magnetic resonance (^1H NMR) stratified by lean and obese individuals reveals a significant difference in the metabolomic profiles of lean EA ($n=5$) and W ($n=5$) individuals ($p=0.006$, $R^2=0.293$, ADONIS) that is not present in obese individuals ($p=0.202$, $R^2=0.141$, ADONIS). (F) Variation in the abundance of the five most abundant phyla in our 16S-seq data reveal significant differences between lean individuals (Firmicutes, Actinobacteria) but not obese individuals between ethnicities. Both lean and obese W individuals had greater Verrucomicrobia compared to their EA counterparts. (G) Volcano plot of ALDEx2 differential abundance testing on genera in stool microbiomes of lean EA versus W individuals, with significantly different ($\text{FDR} < 0.1$) genera highlighted. (H) Abundances of the three genera found to be significantly different between ethnicity: Bacteroides, Blautia, and an unknown genus in the Lachnospiraceae family. (A-D, F-H) $n=22$ EA and $n=24$ W individuals.

Figure 4

Ethnicity-associated differences in the gut microbiota of lean individuals are linked to geographic location. (A) Each symbol represents a subject's ZIP code. Symbols are colored by ethnicity with shape representing lean and obese subjects ($n=44$ subjects). (B) A subset of ZIP Code Tabulation Areas (ZCTAs)

zoomed in to focus on San Francisco are colored by the proportion of each ethnicity (n=27 ZTCAs). The red star indicates a central point (latitude=37.7585102, longitude=-122.4539916) within San Francisco used for distances calculated in (C). (C) Distance to the center of San Francisco, which is indicated by a star in (B), for IDEO subjects stratified by ethnicity and BMI (n=9-13 individuals/group, p-values indicate Wilcoxon rank-sum test). (D) US census data for EA and W residents in ZCTAs from (B) is displayed by ethnic make-up (a total of 489,117 W and 347,200 Asian individuals in these areas). (E-F) Linear regressions of metrics assessing microbial diversity (E) and principal coordinates developed in PCoA analysis (F) against distance of subject's ZIP code to the center of San Francisco. Spearman rank correlation coefficients and p-values are shown for each graph. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 5

Differences in the human gut microbiota between ethnicities are maintained following transplantation to germ-free mice. (A,F) Experimental designs. (A) Germ-free mice fed a LFPF diet received an aliquot of stool from a donor of either ethnicity (per donor n=6 recipient mice, 1 isolator, 2 cages). (F) 5 lean EA and 5 W donors stool microbial communities were transplanted into 20 germ-free recipient mice fed a HFHS diet (per donor n=2 recipient mice, 1 IsoCage). (B,G) Principal coordinate analysis of PhILR Euclidean distances of stool from germ-free recipient mice sampled at (B) six weeks and (G) three weeks post-transplantation. (B) p=0.004, R²=0.630, ADONIS. (G) p=0.001, R²=0.124, ADONIS. Germ-free mice receiving the same donor sample are connected by a dashed line. (C,H) Bacterial richness is significantly higher in mice who received stool samples from lean W donors compared to EA donors. p-values determined using Wilcoxon rank-sum tests. (D,I) Volcano plot of ALDEx2 differential abundance testing on genera in the stool microbiomes between transplant groups. The x-axis represents the fold difference between EA (numerator) and W (denominator) subjects. The y-axis is proportional to the false discovery

rate (FDR). Significantly different genera (FDR < 0.1) are highlighted. (E,J) Abundance of the Bacteroides genus in each experiment with FDR value shown.

Figure 6

Microbiome transplantation of samples from EA and W individuals differentially affects the body composition of genetically identical recipient mice. See experimental designs and group numbers in Figs. 5A,F and donor phenotypic data in Table S8. (A-C, E-G) Percent change in body weight (A,E), fat mass (B, F), and lean mass (C, G) relative to baseline are shown. p-values determined using Wilcoxon rank-sum tests. (D,H) Glucose tolerance test results were not significantly different between groups in either experiment. p-values determined using linear mixed effects models with mouse as a random effect. (A-D) n=6 recipient mice per group. (E-H) n=10 recipient mice per group.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS1.pdf](#)
- [FigS2.pdf](#)
- [FigS3.pdf](#)
- [FigS4.pdf](#)
- [FigS5.pdf](#)

- [FigS6.pdf](#)
- [FigS7.pdf](#)
- [FigS8.pdf](#)
- [IDEOTables.xlsx](#)