The human gut microbiome: methodological issues in characterization and role in trimethylamine *N*-oxide formation

Benjamin C Fu

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Johanna Lampe, Chair

Meredith Hullar

Margaret Madeleine

Program Authorized to Offer Degree:

Public Health – Epidemiology

University of Washington

Abstract

The human gut microbiome: methodological issues in characterization and role in trimethylamine *N*-oxide formation

Benjamin C Fu

Chair of the Supervisory Committee:

Johanna Lampe

Department of Epidemiology

The human gut microbiome has been of great interest in recent years and is being increasingly incorporated into epidemiological research. While early work was primarily interested in characterizing the gut microbiota within healthy individuals, studies have gradually shifted focus towards identifying associations with disease risk and outcomes. The gut microbiome has already been linked to numerous diseases including colorectal cancer and liver disease, as well as allergic and immune diseases. Population-based approaches to studying the microbiome present both challenges and opportunities, including the need to better understand reliability of microbiome samples and the role of microbially-derived metabolites in disease risk.

We used data from the Multiethnic Cohort study (MEC) to address two primary aims: 1) investigate the temporal variability and stability of the fecal microbiome, and 2) identify associations of plasma trimethylamine *N*-oxide (TMAO) and precursors with biomarkers of inflammation and cardiometabolic risk and the fecal microbiome. Temporal variation of the fecal microbiome was assessed within 50 MEC participants who each provided 5 stool samples over

a 2-year period. We calculated the reliability of the overall fecal microbial community using permutational multivariate analysis of variance (PERMANOVA). Taxa and diversity were measured by intraclass correlation coefficients (ICCs). Analyses were additionally stratified based on antibiotics use during the study period. For our second aim, we calculated associations between TMAO and its precursors (choline, carnitine, and betaine) with inflammatory and cardiometabolic risk biomarkers using multivariable regression. Associations between TMAO and the fecal microbiome were also assessed by PERMANOVA and LASSO regression of standardized, centered log-ratio transformed genera.

For our first aim, we found the fecal microbiome to be stable over the two-year study period, with inter-individual variation as the largest source of variation. The majority of microbiome measures were reliable (ICCs>0.40), although genera with very low abundances tended to be variable. Changes in abundances, rather than the complete loss or gain of taxa, were the main source of variation within individuals. Reliability was lower for participants who used antibiotics, although this was largely driven by samples with reported antibiotics use in the month prior to collection.

In our second aim, we identified several associations between TMAO and its precursors with disease biomarkers. In particular, choline indicated risk to adverse health outcomes, while TMAO, choline, carnitine, and betaine were all associated with insulin resistance. Although TMAO explained a small percentage of the overall fecal microbiome variation (<1%), it was associated with several genus-level taxa, including *Desulfovibrio* and two within the family Lachnospiraceae, all of which contain bacteria capable of metabolizing choline into trimethylamine.

In summary, we showed the fecal microbiome to be reliable for use in a population-based study, and when incorporating the fecal microbiome in an epidemiological study, found TMAO to be associated with choline-metabolizing bacteria as well as associations between TMAO and its precursors with inflammatory and cardiometabolic biomarkers.

Table of Contents

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my dissertation chair and advisor, Johanna Lampe, for her incredible mentorship and guidance over the past five years. I am also grateful to the other members of my committee, Meredith Hullar, Margaret Madeleine, and Elhanan Borenstein, for their great insight and help in navigating the dissertation process. I would like to thank Emily White, Tom Vaughan, Steve Schwartz, and Polly Newcomb for their support, as well as Tim Randolph, who was instrumental in the development of my analytical skills and always eager to answer any questions I had. I would also like to thank everyone in the J Lampe lab, along with the Multiethnic Cohort Study investigators and staff, for making all of this research possible. Finally, I want to thank my family and friends for all their encouragement and support during my time in Seattle.

CHAPTER 1
Introduction

Population-based studies of the human microbiome

Using a population approach to study the human microbiome is still a relatively new area of research, with landmark papers by the Human Microbiome Project (HMP) (1) and MetaHIT consortium (2) published less than a decade ago. Although these projects had relatively small sample sizes compared to more traditional epidemiological studies (242 in HMP and 124 in MetaHIT), the decreasing cost of sequencing and a greater recognition of the role of the microbiome in disease risk have led to the establishment of studies with increasingly larger numbers of participants. For example, Goodrich et al. sequenced >1,000 fecal samples from the TwinsUK registry population in a study of the gut microbiome and host genetics (3). A recently published report included data from 1,106 participants in the Belgian Flemish Gut Flora Project, which was validated in a separate Dutch cohort of 1,135 participants (4).

Many important concerns related to epidemiology will need to be addressed in the creation and analysis of these larger population-based studies of the human microbiome. Study design issues such as calculation of sample size and appropriate metadata collection are important and challenging given the multiple methods of characterizing the microbiome. As with population-based studies using other omics platforms, analysis of microbiome data necessitates combining the use of statistical tools that have been traditionally used in epidemiological research with more advanced tools such as machine learning algorithms. Approaches are also being designed for microbiome data specifically that take into account the inherent phylogenetic relationship and data structure.

Quality control of microbiome samples

There are many temporal and technical sources of variation that can impact microbiome samples. Gut microbiome studies often allow for participants to collect stool samples in their own home, thus requiring temporary storage in warmer temperatures (e.g., home freezer and during shipment to study center) than the −80°C typically used for long-term storage. While the effect of storage temperature on collected samples has been assessed in several studies (5-7), how shipping and time between various laboratory processing steps influence microbial community structure is less clear. Most studies addressing this issue keep samples at room temperature or on ice in the laboratory to simulate shipping conditions, although this may not capture fluctuations (e.g. melting) that can occur during transport. Storage time has also been investigated in certain studies (8-10). Many of the quality-control studies that have been published focus on overall community changes, an approach that draws limited conclusions on individual taxa, especially those of lower abundance. There is however, an increasing recognition of the impact of technical variation in microbiome research, and larger efforts such as those by the Microbiome Quality Control (MBQC) project are underway to better quantify these effects as a way to inform investigators developing protocols for studies incorporating microbiome data (11).

Another source of bias in microbiome research is use of internal standards. The benefit of incorporating internal standards during sequencing has been previously shown (12, 13). A study found that processing both positive and negative control samples resulted in many spurious taxa, the presence of which were mainly attributed to wet-lab protocol and contamination from other samples (14). Given the large number of species present in stool samples, determining what should be included in controls is important in order to appropriately normalize counts and identify spurious bacteria.

Temporal variation of the gut microbiome

Some studies have assessed the temporal variation of the microbiome in the gut and other anatomical sites, although these have tended to have small sample sizes, variable sampling periods, or were conducted for shorter periods of time. Costello et al. collected samples from multiple body sites including the gut, mouth, nares, hair, and skin both one day and 3 months apart (15), finding significantly less variation within individuals than between individuals, and greater variation over 3 months than 24 hours. Gut and skin exhibited greater variability across participants than oral samples. Several other studies have also shown there to be greater variation between individuals than within individuals over time (16, 17). Skin has also been found to be more variable in terms of the number of different taxa present when compared to other body habitats (18). Although microbial communities as a whole have been found to be stable over time, specific taxa may have more temporal variability. In a study of 37 healthy adults, 60% of bacterial strains remained after 5 years, while extrapolated data indicated that a majority of strains would be expected to remain after several decades (19).

Additional studies are important for measures of the microbiome obtained from other sequencing platforms. Mehta et al. analyzed metagenomic and metatranscriptomic sequence data from stool samples of 308 men within the Health Professionals Follow-up Study, finding relatively higher levels of inter-individual variation for taxonomic and functional measures, whereas within and between-individual variation for metatranscriptomic profiles were much more similar (20). Only 0.79% of transcripts had intraclass correlation coefficients greater than 0.40 in that study, suggesting metatranscriptomic profiles to be much more dynamic.

Trimethylamine *N*-oxide and disease risk

Gut microbes produce a wide range of compounds through the metabolism of dietary components. Among the most studied are short chain fatty acids (SCFAs), which are generated from fermentation of dietary fiber and serve important functions, including as a major energy source for colonocytes and involvement in regulating many pathways related to energy

metabolism (21). The gut microbiota can synthesize various other beneficial compounds including vitamins and essential amino acids, some of which can only be produced by bacteria and archaea (22-24).

However, the gut microbiome can also generate compounds that have been proposed to have harmful effects, including trimethylamine *N*-oxide (TMAO), which has recently been implicated in several diseases, most notably through a series of cardiovascular disease (CVD) studies out of Cleveland Clinic (Cleveland, OH). In a cohort study of adults undergoing elective diagnostic cardiac catheterization, those in the highest quartile of plasma TMAO concentration had significantly increased risk of developing a major adverse cardiovascular event (myocardial infarction, death, or stroke) over three years compared to those in the lowest quartile (adjusted HR=1.43, 95% CI: 1.05-1.94) (25). TMAO was also prognostic in subgroups with lower risk of major adverse cardiovascular events (e.g. no hypertension, no history of coronary artery disease, diabetes mellitus) (25, 26). Dietary choline supplementation of Apolipoprotein E knockout ($ApoE^{-/-}$) mice led to increased TMAO levels and increased levels of pro-atherogenic markers including foam cell formation and scavenger CD36 and SR-A1 surface protein levels in macrophages (27). Evidence of TMAO as the causative agent was strengthened by the lack of a proatherogenic effect in mice on a similar diet given broad-spectrum antibiotics (27). Dietary supplementation of $ApoE^{-/-}$ mice with L-carnitine led to increased TMAO and atherosclerotic plaque, inhibition of reverse cholesterol transport, and reduced expression of bile acid transport genes in the liver (28).

Subsequent studies have found relationships between TMAO and several other chronic diseases. Plasma TMAO was associated with rectal cancer (OR=3.38, 95% CI: 1.25-9.16) and overall colorectal cancer among women with below median plasma levels of vitamin B12 (OR=2.44, 95% CI: 1.59-3.75) in a nested case-control study within the Women's Health Initiative (29). In a prospective metabolomic analysis, TMAO was elevated in prostate cancer

cases compared to matched controls (OR=1.36, 95% CI: 1.02-1.81) (30). Plasma TMAO showed a linear dose-response relationship with type 2 diabetes in a case-control study based in China, with more than two times the odds of diabetes (OR=2.55, 95% CI: 1.99-3.28) for those in the fourth quartile of TMAO compared to the first quartile (31). Higher TMAO concentrations have also been found to be associated with heart failure and kidney diseases (32-35).

Potential role of TMAO in systemic inflammation and cardiometabolic risk

Potential mechanisms of action of TMAO in relation to diseases besides CVD are unclear. Given the central role of inflammation in all phases of the atherosclerotic process (36) and its association with many chronic diseases, TMAO has been hypothesized to influence disease risk through inflammation-related pathways. Many of the human studies of TMAO and inflammation to date have been limited by small numbers or the study of unhealthy individuals, and have shown mixed results. Consumption of eggs, a major source of phosphatidylcholine, increased TMAO concentrations but did not change CRP levels in a dose-response study of six healthy individuals (37) TMAO was also not associated with CRP in observational studies of patients with heart failure (33) or chronic kidney disease (CKD) (35) but was inversely associated in hemodialysis patients (38). Mice given TMAO had increased expression of the pro-inflammatory cytokine MCP-1 and decreased expression of the anti-inflammatory cytokine IL-10 in adipose tissue, suggesting TMAO may influence inflammatory biomarkers other than CRP (39). This is supported in a recent study of healthy adults where TMAO was associated with higher concentrations of TNF-α and soluble TNF receptors (40). Another study found that treatment of human aortic endothelial cells and human vascular smooth muscle cells with TMAO increased expression of several inflammatory biomarkers, although pretreatment of these cells with an NF-κB inhibitor was able to block the TMAO-induced expression (41).

Animal studies suggest a role of TMAO in diabetes and insulin resistance, as supplementing mice with TMAO increased fasting insulin levels and impaired glucose tolerance

(39) while mice with impaired growth hormone receptor signaling produced higher concentrations of TMA and TMAO (42). Additionally, insulin has been shown to suppress FMO3 activity, and knockdown of FMO3 led to the suppression of FoxO1 and improved glucose tolerance in insulin-resistant mice (43). There is also growing evidence in population-based studies, as Shan et al. (31) recently identified a positive association between TMAO and type 2 diabetes in a case-control study.

There is also epidemiological evidence that choline and betaine contribute to cardiometabolic risk. Konstantinova and colleagues reported that an unfavorable cardiovascular risk profile was associated with higher levels of plasma choline and lower levels of plasma betaine (44). Similar results were found in a study of older US adults (45). A third study also had similar results for betaine, which was negatively associated with BMI, non-high-density lipoprotein cholesterol, and triglycerides (46). Complementary studies of TMAO and cardiometabolic biomarkers would be useful in better understanding its potential disease mechanisms and for helping to shed light on previously shown associations with cardiovascular disease in particular.

Bacteria involved in TMAO formation

TMAO is formed by a two-step process in humans. Gut bacteria metabolize dietary quaternary ammonium compounds (e.g. choline, carnitine, and betaine) to trimethylamine (TMA), which is rapidly oxidized by human hepatic enzymes to form TMAO (47, 48). Early studies of microbial TMA formation were conducted using culture-based techniques and often tested a single species or a group of similar species at a time. Several bacteria have been found to produce TMA on a choline-based medium, including *Streptococcus sanguinis, Escheria coli, Aerobacter aerogenes, Desulfovibrio desulfuricans, Proteus rettgeri,* and *Shigella alkalescens* (49, 50). The bacterial species *Serratia marcescens* and *Acinetobacter calcoaceticus* can produce TMA from carnitine and γ-butyrobetaine (51, 52). Administration of a broad-spectrum

antibiotic in mice led to near complete suppression of detectable TMAO after a high choline or L-carnitine diet (27, 28), further highlighting the importance of microbes in this process. One 16S rRNA-based analysis of human microbiota found associations between several microbial groups and TMAO concentrations, but only included genus-level comparisons and 53 participants (28). Another study individually inoculated 79 human gut isolates in a diluted gut medium with choline, identifying 8 species (*Anaerococcus hydrogenalis, Clostridium asparagiforme, Clostridium hathewayi, Clostridium sporogenes, Escherichia fergusonii, Proteus penneri, Providencia rettgeri*, and *Edwardsiella tarda*) that were able to produce TMA, all of which belonged to the phyla *Firmicutes* and *Proteobacteria* (53). Colonization of germ free mice with these bacteria also resulted in detectable TMAO concentrations in serum.

Microbial functional genes involved in TMAO formation

Although it is known that TMA is microbially produced by cleavage of the C-N bond in its precursors, the microbial genetic and biochemical mechanisms involved have only been recently investigated using bioinformatics techniques. These studies have typically searched within TMA-producing bacteria either for homologs of genes involved in metabolism of structurally related compounds (54) or genes neighboring those involved in metabolism of TMA precursors (55). To date, two gene clusters have been identified. The choline utilization (*cut*) gene cluster in *Desulfovibrio desulfuricans* encodes a glycyl radical enzyme (CutC) and a glycyl radical protein (CutD). Another gene cluster in *Acinetobacter baumannii* encodes a two-component Rieske-type oxygenase/reductase (cntA, cntB) that catabolizes choline and carnitine, respectively (54, 55). Mutagenesis of either or both genes in these clusters greatly reduced TMA formation. Microbes without these genes have been found to be able to produce TMA, suggesting other genes may be involved (53).

Subsequent studies have used bioinformatics approaches to identify bacteria that possess these enzymes (i.e. have the potential to form TMA) by searching through whole

metagenomics shotgun sequencing data. Bacteria identified through these methods appear to be restricted to the phyla *Actinobacteri, Firmicutes,* and *Proteobacteria* (56). Although TMA-forming bacteria are present in fecal samples, they represent a small percentage of the overall microbial community. The *cut* gene cluster also appears to be more prevalence than the *cnt* cluster (56).

Dissertation aims

In this dissertation, we used data from the Multiethnic Cohort study (MEC) to address two aims: 1) investigate the temporal variability and stability of the fecal microbiome, and 2) identify associations of plasma TMAO and its precursors with biomarkers of inflammation and cardiometabolic risk to inform our understanding of the role of the fecal microbiome in metabolism. The results from testing these aims will help provide guidance in developing sample collection protocols for longitudinal microbiome research, as well as a better understanding of the gut microbiome's role in disease risk.

CHAPTER 2

Temporal variability and stability of the fecal microbiome: the Multiethnic Cohort Study

ABSTRACT

Background: Measurement reliability and biological stability need to be considered when developing sampling protocols for population-based fecal microbiome studies.

Methods: Stool samples were collected biannually over a two-year period and sequenced for the V1-V3 region of the 16S rRNA gene in 50 participants from the Multiethnic Cohort Study. We evaluated the temporal stability of the fecal microbiome on a community level with permutational multivariate analysis of variance (PERMANOVA), as well as on taxa and diversity measures with intraclass correlation coefficients.

Results: Inter-individual differences were the predominant source of fecal microbiome variation, and variation within individual was driven more by changing abundances than the complete loss or introduction of taxa. Phyla and diversity measures were reliable over the two years. Most genera were stable over time, although those with low abundances tended to be more dynamic. Reliability was lower among participants who used antibiotics, with the greatest difference seen in samples taken within one month of reported use.

Conclusions: The fecal microbiome as a whole is stable over a two-year period, although certain taxa may exhibit more temporal variability.

Impact: When designing large epidemiologic studies, a single sample is sufficient to capture the majority of the variation in the fecal microbiome from 16S rRNA gene sequencing, while multiple samples may be needed for rare or less abundant taxa.

Introduction

In the past decade, the gut microbiome has been of great interest in health research, with diseases such as colon cancer (57), inflammatory bowel disease (58), and cardiovascular disease (27) already linked to both community-wide shifts and changes in specific bacterial taxa. Our ability to identify associations between gut microbes and diseases will be greatly improved with the continued establishment of well-powered population-based longitudinal studies coupled with the decreasing costs of DNA sequencing. In order to conduct these large-scale studies, standardized methods that provide reliable estimates need to be implemented. Several studies have already investigated technical sources of variability due to different aspects of sample collection (59, 60), processing (11, 61), and sequencing (62-64).

Having reliable estimates that sufficiently capture temporal variation of the gut microbiome is also crucial**.** Microbial communities are complex and constantly changing in response to their environment. Factors such as diet (65-68), use of antibiotics and other medications (4, 69), and exposure to pathogens (70) can have a pronounced impact on bacteria residing in the gut and other anatomical sites. In the context of epidemiologic research, a microbiome with dramatic fluctuations over time could require multiple sample collections or increased sample sizes for longitudinal studies. Previous studies have evaluated variation of the fecal microbiome over time, but have involved small numbers of participants (15, 16), variable sampling periods (19), or only Caucasian populations (17), which may limit generalizability. Here, we assessed the temporal variability of the fecal microbiome in 50 older adults from a multiethnic population with biannual sampling over a two-year period.

Methods

Study participants

The Multiethnic Cohort study (MEC) is a prospective cohort study conducted in Hawaii and Los Angeles County that was designed to investigate the association of lifestyle and genetic

factors with the incidence of cancer and other chronic diseases. The study design, recruitment, and baseline characteristics have been described previously (71). Briefly, 215,251 men and women between the ages of 45-75 from primarily five racial/ethnic groups (African-American, Japanese-American, Latino, Native Hawaiian, and white) were enrolled into the study from 1993-1996 by completing a self-administered 26-page mailed questionnaire. Over 1800 of these participants (aged 60-77) were recruited in 2013-2017 as part of the MEC Adiposity Phenotype Study (APS) to investigate the relationships between the exposome, genome, microbiome, and metabolome with body fat distribution. Exclusion criteria for the MEC-APS included reported BMI outside the range of 18.5-40 kg/m^2; oral or injection antibiotic use in the past 3 months; current or recent (<2 years) smoking; flu shot or other vaccinations in the past month; substantial weight change (>20 lbs) in the past 6 months; soft or metal implants; ileostomy or colectomy; dialysis; insulin or thyroid medication; and any of the following procedures or treatments in the past 6 months: chemotherapy, radiation therapy, corticosteroid hormones, prescription weight-loss drugs, endoscopy or irrigation of the large intestine. Percent body fat was measured by whole-body dual-energy X-ray absorptiometry (DXA) scans (72). Fifty individuals were randomly selected from the APS participants to have an equal distribution by sex (25 male and 25 female), the five main ethnic groups within the MEC (10 African American, 10 Japanese American, 10 Native Hawaiian, 10 Latino, and 10 whites), and BMI categories (within each sex-ethnic group, one from each of 22-24.9, 25-26.9, 27-29.9, 30-34.9 kg/m^2 and one either from 18.5-21.9 or 35-40 kg/m^2) in which to conduct our longitudinal fecal microbiome study. Institutional Review Board approval was obtained from all participating institutions and informed written consent was obtained from the study participants.

Sample collection

Over a two-year period, each participant was asked to collect a stool sample once every six months for a total of five samples. Stool samples were collected at home using a collection

tube containing 5 mL RNAlater (Fisher Scientific) and sterile 5 mm glass beads (Ambion) to facilitate sample dispersion in RNAlater. Samples were then frozen overnight and either brought in or mailed to the study clinic the following morning. Collection materials and procedures have been described in detail previously (73). Along with each sample, participants were asked to fill out a stool collection questionnaire that included items on collection time, special diets, and consumption of probiotic foods in the past six months. The questionnaire also asked whether participants were treated with an oral, injection, or IV form of antibiotics in the past six months, and the most recent month antibiotics were taken. If, at baseline, the participants reported to have received antibiotic therapy during the past six months, collection was deferred by six months and the baseline eligibility questionnaire was re-administered.

Sample processing

Stool samples were shipped on dry ice from study centers in Honolulu, HI and Los Angeles, CA to the Fred Hutchinson Cancer Research Center (FHCRC) in Seattle, WA. Stool samples collected in RNAlater were thawed and homogenized at 10,000 RPM on ice for 30 seconds (Omni Tissue Homogenizer, Omni International, GA). Homogenized sample (300 µL) was transferred into four FastPrep tubes (MP Biomedical, Santa Ana CA) along with 0.3 g zirconium beads (Biospec Products, Bartleville OK) which were previously sterilized in an oven (180°C for >2 hours), and stored at -80 °C. For DNA extraction, two FastPrep tubes from each sample were thawed on ice. Sterile phosphate buffered saline (300 µL) was added to each of the tubes, which were then centrifuged at 14,000 RPM for 10 minutes. The supernatant was removed and discarded. Preheated ASL buffer (50 ^{o}C; 700 µL; QIAGEN, Germantown MD) was added to the pellet in each sample tube. FastPrep tubes were placed in a FastPrep bead beater 24-5G (MP Biomedical) at 5.5 m/s for 45 seconds, followed by 95°C (Thermomixer, Eppendorf, Hauppauge NY) for 15 minutes at 15,000 RPM, and centrifuged for 3 minutes at 15,000 RPM. 520 µL of the supernatant was placed in a 1.5 mL tube containing an InhibitEX tablet (QIAGEN).

Eppendorf tubes were centrifuged for 3 minutes at 15,000 RPM. The remaining DNA extraction procedures followed the standard QIAcube protocol for human stool (Qiagen). Final elution of DNA was performed with 200 µL elution buffer (AE buffer; QIAGEN). DNA concentrations and purity were determined using the NanoDrop 8000 Spectophotometer (ThermoFisher Scientific, Waltham, MA) and gel electrophoresis. Working stocks were diluted in AE buffer (QIAGEN) from genomic DNA and samples were stored at -20°C until shipped for sequencing.

Samples for sequencing were prepared using a working stock at final concentration of 20 ng/µL. Samples from the same participant were processed together in the same batch. FHC samples were used to assess variation in library preparation and sequencing batches. FHC samples were prepared by pooling stool from 6 participants outside the time-series study who had not used antibiotic in the past three months. From each participant, we collected five tubes of stool, with each tube containing 5 mL RNAlater and two scoops of stool that were stored at -80°C. All five tubes from each participant were thawed on ice, briefly homogenized individually, and then all combined into one container. Homogenized stools (400-500 µL) were distributed into multiple aliquots in FastPrep tubes and stored at -80°C. To assess DNA extraction, we used duplicate stool samples from three individuals outside of the time-series study who had not used antibiotics in the past three months. Two stool samples per individual were collected in RNAlater and frozen at -80 ^{o}C for one week. Samples were thawed on ice, homogenized and extracted using the protocol outlined above. Intraclass correlation coefficients (ICCs) for extraction duplicates were ≥0.93 for alpha diversity measures, ≥0.99 for the first PCoA axis for unweighted and weighted UniFrac, and ≥0.97 for the four most abundant phyla.

For paired-end sequencing of the V1-V3 region, the 27F mod forward PCR primer sequence was 5'-AGRGTTNGATCMTGGCTYAG-3'. The 519R reverse PCR primer sequence was 5'- GTNTTACNGCGGCKGCTG-3'. A 25-cycle PCR was performed using the HotStarTaq Plus Master Mix Kit (Qiagen, USA) under the following conditions: 94°C for 3 minutes, followed by 28 cycles of 94°C for 30 seconds, 53°C for 40 seconds, and 72°C for 1 minute, after which a

final elongation step at 72°C for 5 minutes was performed. After amplification, PCR products were checked in 2% agarose gel to determine the success of amplification and the relative intensity of bands. Multiple PCR products were pooled together in equal proportions based on their molecular weight and DNA concentrations. Pooled samples were purified using calibrated Ampure XP beads (Beckman Coulter, USA). The pooled and purified PCR products were used to prepare the Illumina DNA library using a ligation process (TruSeq Nano DNA LT, QIAGEN) which included Illumina adapters, pads, linkers and an 8 base pair (bp) barcode index. Sequencing was performed on the MiSeq using MiSeq Reagent Kit v3 following the manufacturer's guidelines to obtain 2x300 bp paired-end reads (Illumina, San Diego, CA). FastQ files were exported and securely transferred (BaseSpace, Illumina) to FHCRC for bioinformatic analysis.

Microbiome bioinformatic data processing

To classify bacterial taxonomy, sequences were processed using QIIME v.1.8 (74). Sequences were joined with the fastq-join method, using min_overlap=15 and perc_max_diff=12. Sequences were filtered with split_libraries_fastq.py with q parameter set to 25, and defaults otherwise. The Nelson two-step method was used for operational taxonomic unit (OTU) generation at 97% similarity using the SILVA database (release 111, clustered at the 97% similarity level) for closed reference OTU picking following the UCLUST algorithm (75). The OTU table was filtered using the QIIME script filter_otus_from_otu_table.py with --min_count_fraction set to 0.00005 as recommended in Navas-Molina et al (76). Additional OTU entries were filtered out if they were detected as chimeras using QIIME's identify_chimeric_seqs.py script with the blast_fragments method (77). The sequences were classified using the matching SILVA taxonomy for OTUs found in the first step of the Nelson method, and MOTHUR's naive Bayesian Classifier (78, 79) trained against the SILVA database (release 111, clustered at the 97% similarity level) for OTUs found in the second step.

Sequences were aligned to the SILVA 16S rRNA gene reference alignment (80) using the NAST algorithm (81). Sequences that did not align to the appropriate 16S rRNA gene region were removed. The phylogenetic tree was constructed following the FastTree method (82). Sequence counts for each sample ranging from phylum to genus level were generated without rarefaction. Alpha diversity measures [phylogenetic diversity (83), Shannon index (84), Chao1 index (85)] and beta diversity matrices [unweighted and weighted UniFrac (86, 87)] were calculated in QIIME based on the average of 10 sub-samples with rarefaction to 10,000 sequences per sample.

Statistical analysis

Differences in fecal microbiota composition were assessed using two phylogenetic beta diversity metrics, unweighted UniFrac and weighted UniFrac. Unweighted UniFrac is a qualitative measure that captures differences in the presence and absence of OTUs, while weighted UniFrac is a quantitative measure that additionally incorporates information on the relative abundance of OTUs (87). Principal coordinate analysis (PCoA) plots using the first two PCoA axes were generated for both unweighted and weighted UniFrac distances using the 'cmdscale' function in R. The variation in microbial community structure explained by individual, time point, sample receipt time, and antibiotic use was determined by PERMANOVA (999 permutations) for both unweighted and weighted UniFrac distances using the 'adonis' function from the R package 'vegan' (88). Due to the prevalence of use and impact of antibiotics, we stratified our analyses based on whether participants reported any antibiotic use during the 2-year study period.

To determine whether samples more closely resembled other samples from the same individual or samples from different individuals, we matched each non-baseline sample with the baseline sample it was most similar to as defined by the shortest distance using unweighted and

weighted UniFrac metrics. We determined whether each pair of samples belonged to the same individual and then calculated the proportion of pairs that both belonged to the same person.

Taxon abundances are often normalized by converting raw counts into relative abundances per sample. Although this addresses the issue of varying sequencing depth, the subsequent data are constrained to a simplex due to the unit-sum constraint and, while useful for characterization, may not be appropriate for use with standard statistical approaches. Here, we applied the interquartile log-ratio transformation (IQLR) for all taxa abundances, which allows for analysis of compositional data by calculating log-ratios of abundances and has been shown to be effective in producing approximately multivariate normal data (89, 90).

We used ICCs to assess the reliability of several commonly used microbiome measures, including the four most abundant phyla (Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria), the three alpha diversity measures described above, and three beta diversity measures (first PCoA axis for unweighted UniFrac, weighted UniFrac, Bray-Curtis, and Jaccard). ICCs were calculated by first fitting a linear mixed effects model with a random effect for participant using the 'lmer' function in the R package lme4 (91), and then dividing the between-individual variation by the total variation from the model using the 'icc' function in the R package sjstats. To assess the reliability of genera, we computed ICCs for abundances as well as presence/absence. ICCs for abundances were calculated using the approach described above. ICCs for presence/absence were calculated by first converting the genus abundance table to a presence/absence table by replacing all counts greater than 0 with 1. Any genus that was present in every sample was excluded as the ICC would be undefined due to no variation. Then, a generalized linear mixed effects model with a binomial distribution and a random effect for participant was fitted using the 'glmer' function in lme4. ICCs were then computed using the 'icc' function in sjstats. For all ICC measures, reliability was considered excellent for ICC≥0.75, good for 0.74≥ICC≥0.60, fair for 0.59≥ICC≥0.40, and poor for ICC≤0.39 (92).

Next, we assessed whether one sample is sufficient or multiple samples over time are necessary when using the fecal microbiome in an association analysis with a health-related outcome. Since multiple studies have aimed to link the fecal microbiome with obesity (93-95), we selected baseline percent body fat as a benchmark for assessing stability of the association over time. The variation in fecal microbiome composition explained by baseline percent body fat was calculated using PERMANOVA R^2 at baseline, as well as with the addition of subsequent samples (e.g. including baseline and the 6-month sample) using unweighted and weighted UniFrac to examine differences in the association when incorporating multiple time points.

We also explored recovery of the fecal microbiome from antibiotics among participants who reported antibiotic use during the two-year study period. Participants were excluded from this analysis if they did not provide the last date of antibiotic use, or had a baseline sample that failed laboratory quality control. We assessed recovery in samples that were taken after the first reported use of antibiotics only (i.e. samples were not included if they were collected after two or more courses of antibiotics), and categorized them based on time between last antibiotic use and date of sample collection (0-1 months, 1-3 months, 3-6 months, 6-12 months, 12-24 months). Percent change in alpha diversity (Shannon index and phylogenetic diversity) was calculated by dividing the difference in diversity between a sample and the baseline sample by the diversity of the baseline sample, and multiplying by 100. We also assessed changes in beta diversity by using unweighted and weighted UniFrac distances between each sample and the baseline sample corresponding to the same individual. Differences between the six-month and baseline samples for those not reporting antibiotic use were also included as a comparison. Changes in alpha diversity for each time interval was assessed using a one-sample t-test for a mean of zero. All analyses were conducted in R version 3.4.3.

Results

Participant characteristics

Participants (n=50) had a mean ± SD age of 68.6 ± 2.7 years, and were equally distributed by sex and the five race-ethnic groups of the MEC (Table 1) in accordance with our recruitment strategy. 23 (46%) participants reported using antibiotics at least once during the study period. Those who took antibiotics were more likely to be male and Latino. Samples were collected 186.8 ± 36.0 (mean ± SD) days apart. Alpha diversity and phyla abundances were comparable between antibiotics use group (Supplemental Table 1).

Temporal variation of the fecal microbiome

In our samples, we identified 10 phyla, 20 classes, 26 orders, 46 families, 93 genera, and 1220 OTUs. Three genera were present in every sample (Bacteroides and two belonging to Lachnospiraceae). There was 38,768 ± 11,596 (mean ± SD) sequences per sample and an average sequence length of 493 bp. ± 5 bp (mean ± SD). From the PCoA plots based on unweighted UniFrac, samples from the same individual generally clustered together (Figure 1A-C), particularly among those who did not take antibiotics (Figure 1A). There was greater overlap of samples between individuals when using weighted UniFrac (Figure 1D-F). The majority of microbiome variation was due to inter-individual differences (Table 2), accounting for 70% and 78% of the total unweighted UniFrac variation for those on and not on antibiotics, respectively. Inter-individual variation explained slightly less but remained the largest source of variation when using weighted UniFrac, accounting for 66% and 70% of the variation for those on and not on antibiotics, respectively. Variation was explained minimally by sample time point and days to receipt at study center (<1%). Inter-individual differences and antibiotic use were significant sources of microbiome variation while sample time point and days to receipt at study center were not.

To assess whether a single sample was representative of an individual's microbiome over time, we matched each non-baseline sample to the baseline sample with the shortest UniFrac distance. The majority of non-baseline samples matched to the baseline sample of the

same individual when using unweighted UniFrac distances (no antibiotics: 83%; antibiotics: 72%). Fewer samples matched correctly for weighted UniFrac, with about one-third of samples being most similar to the same person's baseline sample (no antibiotics: 34%; antibiotics: 30%).

Reliability of microbiome measures

We next assessed fecal microbiome variability over time using ICCs of taxa and diversity measures. Among all participants, the four phyla had fair reliability, with ICCs between 0.56-0.59 (Table 3). Differences were seen when stratifying by antibiotic use, as ICCs were consistently higher among those not taking antibiotics compared to those who did. The majority of genera had at least fair reliability. For abundance measures, 79% of genera in the no-antibiotics group and 74% in the antibiotics group had ICC>0.40 (Figure 2A-C). For presence/absence, 86% in the no-antibiotics group and 84% in the antibiotics group had ICC>0.40 (Figure 2D-F). Genera with poor reliability were typically those with low abundance or low prevalence (Supplemental Table 2).

Alpha diversity measures tended to have better reproducibility than individual phyla measures in the no-antibiotics group. Phylogenetic diversity had the highest reproducibility, followed by Shannon diversity and the Chao1 estimator. The ICCs of all three alpha diversity measures were greater in the no-antibiotics group than in the antibiotics group. We also assessed beta diversity reproducibility, finding unweighted UniFrac PC1, Bray-Curtis PC1, and Jaccard PC1 to have excellent stability over time regardless of antibiotic use, and weighted UniFrac PC1 to have good stability over time in both groups (Table 3).

Microbiome-body fat associations

To test the effect of variation in the microbiome over time on a relevant health outcome, we modeled the microbiome-body fat association starting with the baseline sample, followed by the addition of subsequent samples. Percent body fat had a wider range of values in the

antibiotics group (11.9%-46.8%) than in the no-antibiotics group (21.4%-50.3%). The variation of the fecal microbiome explained by percent body fat did not fluctuate with the addition of subsequent samples, and remained relatively stable (Supplemental Table 3) whether using all participants (0.029-0.034), only participants not on antibiotics (0.035-0.056), or only participants on antibiotics (0.056-0.070) for unweighted UniFrac. Weighted UniFrac measures were slightly more variable but still consistent over time.

Recovery from antibiotics use

We also explored microbiome recovery from antibiotics by comparing post-antibiotic use samples to the pre-antibiotic baseline sample among participants who took antibiotics. Although none of the time intervals were significantly different from zero for Shannon index (Figure 4A), changes in the first month were the most variable. The percent change in phylogenetic diversity for samples taken in the first month were significant (Figure 4B). Box plots for beta diversity suggested little difference in unweighted UniFrac distance compared to baseline samples across time intervals, but differences were larger overall among antibiotic users than participants not taking antibiotics (Figure 4C). Recovery over time was more evident for weighted UniFrac, with distances from baseline after the first month post-antibiotics closely resembling those not taking antibiotics (Figure 4D)

Discussion

Using several approaches, we showed the fecal microbiome as a whole to be relatively stable over a two-year period. Samples from the same participant clustered together and an association analysis between the overall community structure and baseline body fat showed consistent results throughout the study period. Much of the variation was due to changes in taxa abundances rather than the complete loss or gain of taxa. Although reliability among

participants who reported antibiotic use tended to be lower than among those who did not, the largest differences appeared to be among samples taken within a month of antibiotic use.

As with previous studies (15-17), inter-individual differences were the dominant source of variation, as evident in the wide phylum distribution of our samples, with relative abundances ranging from 15.8%-89.6% for Firmicutes and 6.7%-67.0% for Bacteroidetes. When matching non-baseline samples to the most similar baseline sample, we found that the majority matched to the same individual when using unweighted UniFrac, while fewer matched using weighted UniFrac, suggesting that changes over time were driven more by changing abundances of taxa rather than their presence or absence. Claesson et al. (17) conducted a similar analysis on stool samples collected 3 months apart. Although they found less discrepancy between weighted and unweighted UniFrac measures compared to our results, there was a similar pattern where fewer samples were matched when using weighted UniFrac.

There is also growing interest in studying the associations of specific taxa with disease, such as *Fusobacterium* and colon cancer (96, 97) or Christensenellaceae and obesity (3). We found phyla measures, as well as the majority of genera, to be reliable. Temporal variation was more of a concern for genera with very low abundances or prevalence, some of which could be taxa that are transient and not representative of an individual's microbiome over time or those that are near the detection limit and are thus not able to be consistently identified. Larger sample sizes may be necessary if these are of particular interest to a study. Less abundant taxa exhibited lower reproducibility in other methodologic studies as well (60).

In a set of exploratory analyses, we were able to assess the ability of the fecal microbiome to recover from antibiotics. Antibiotic use explained a small but significant proportion of the overall fecal microbiome variation. The strongest and most variable effect on diversity generally occurred in the weeks following use, with samples more closely resembling pre-antibiotic levels in the months that followed. Sampling the fecal microbiome one year (98), and even four weeks after antibiotics (99), has shown return in alpha diversity to pre-treatment

levels, although recovery varied for different taxa. Similarly, the ELDERMET study reported that alpha diversity among those who reported antibiotic use within the past month was not significantly different from those who did not (100). However, nine genera were found to be different when using 16S rRNA gene sequencing, as were *Bifidobacterium* levels when measured by culture. With frequent sampling, Dethlefsen et al. was able to show that adults undergoing courses of ciprofloxacin saw decreases in OTU richness, phylogenetic diversity, and Shannon index within 3-4 days of administration (69). Participants began to recover within a week after taking the antibiotic, although the time needed to reach a stable level varied among participants and alterations in the abundances of certain taxa were apparent. While the gut microbial community as a whole may be able to recover from antibiotics, lingering effects on specific taxa highlight the need for the development of antibiotics with more targeted effects as an alternative to those that act on a broad range of bacteria.

Antibiotic use has also been associated with disease risk factors, including body weight, in animal models and epidemiologic studies, (101). We found that the fecal microbiome explained more variation in body fat among individuals who used antibiotics. Assuming that antibiotic use captured in our study reflects use before baseline (when percent body fat was measured), this finding might suggest a greater contribution by the altered microbial community structure to metabolic regulation and energy homeostasis. There is evidence that the type of antibiotic may have a different effect on overweight and obesity as well, since a longitudinal study by Bailey et al. (102) found an association between early-life exposure to broad-spectrum antibiotics with obesity, but not for narrow-spectrum antibiotics.

Our study design had several strengths. Among studies on temporal variation of the fecal microbiome, ours has one of the most ethnically diverse populations to date and one of few using elderly participants. We were able to collect samples on a consistent schedule over a longer period than other population-based studies, which typically collect samples for only a few

months. The retention rate of the 50 enrolled study participants over the two years was also very high, with 49 sending all five samples and 1 participant only missing one.

A limitation of our study was that we were not able to assess other sources of microbiome variation, such as travel or recent health, as these were not included in the questionnaire that was filled out at each stool collection. Our study also used 16S rRNA gene data. Additional studies measuring temporal variation of other aspects of the gut microbiome, such as the metagenome and metatranscriptome (20), are crucial. Another limitation was that we did not have information on what types of antibiotics were used or reason for use. Antibiotics have varying mechanisms of action that include targeting bacterial cell walls or membranes, protein synthesis, and DNA or RNA synthesis (103). A two-center randomized controlled trial in the United Kingdom and Sweden reported different responses to the Shannon index from four different antibiotics, with effects on the gut microbiome ranging from no difference after one week to sustained reduction at one year (104). As antibiotics are frequently prescribed for treating a variety of infections, as well as for prophylaxis in preventing infections among high-risk patients, the disease state may also modify the effect of antibiotic treatment. However, the temporal trend we saw was comparable to studies conducted in participants who were healthy at the time of antibiotic administration (69, 99).

In summary, we showed that a single assessment sufficiently captures the majority of fecal microbiome measurements in a population-based study, but special consideration should be taken with very rare or low abundant taxa. The assessment of methodologic issues, such as our test of the reliability of measurements, is an important step in designing robust, effective population-based studies to evaluate the role of the fecal microbiome in disease risk.

Acknowledgements

We thank the Multiethnic Cohort Study participants who generously donated their time and effort to the Adiposity Phenotype Study.

We acknowledge the contribution of the study staff members whose excellent performance made this research possible: the Recruitment and Data Collection Core staff at USC (Adelaida Irimian, Chanthel Figueroa, Brenda Figueroa, Karla Soriano) and UH (Dr. Terrilea Burnett, Naomi Hee, Clara Richards, Cheryl Toyofuku, Hui Chang, Janice Nako-Piburn); the Data Management and Analysis Core staff at USC (Zhihan Huang) and UH (Maj Earle, Joel Julian, Anne Tome); the Project Administrative Core staff at UH (Eugene Okiyama); the Laboratory staff at FHCRC (Orsalem Kahsai, Wendy Thomas, Elizabeth Traylor, and Crystal Voyce); and Bioinformatics staff at FHCRC (Keith Curtis).

Table 1. Characteristics of MEC study participants by any reported antibiotic use.

	No antibiotic use (n=27)	**Antibiotic use (n=23)**	**Total (n=50)**
Age, years	68.2 ± 2.9	69.0 ± 2.3	68.6 ± 2.7
Female	15 (55.6)	10 (43.5)	25 (50)
Race/ethnicity			
African American	5 (18.5)	5 (21.7)	10 (20)
Japanese American	6 (22.2)	4 (17.4)	10 (20)
Native Hawaiian	7 (25.9)	3 (13.0)	10 (20)
Latino	3 (11.1)	7 (30.5)	10 (20)
White	6 (22.2)	4 (17.4)	10 (20)
Education, years	15.0 ± 2.5	13.8 ± 3.4	14.4 ± 3.0
Smoking status			
Never	19 (70.4)	17 (73.9)	36 (72.0)
Former	8 (29.6)	6 (26.1)	14 (28.0)
Body fat %	33.2 ± 6.8	32.2 ± 9.2	32.8 ± 7.9

Mean ± SD for continuous variables and n (%) for categorical variables

Table 2. Microbiome variation explained by inter-individual differences, time point of sample, days to sample receipt at study clinic, and antibiotic use calculated using a distance-based analysis of variance.

	No antibiotic use (n=27)		Antibiotic use (n=23)		Total (n=50)	
	R^2	p	R^2	p	R^2	p
Unweighted UNIFRAC						
Individual	0.777	<0.001	0.696	<0.001	0.743	<0.001
Time point	0.006	0.797	0.008	0.643	0.004	0.424
Days to receipt	0.007	0.640	0.012	0.180	0.005	0.178
Antibiotic use					0.017	0.001
Weighted UNIFRAC						
Individual	0.701	<0.001	0.655	<0.001	0.687	<0.001
Time point	0.003	0.909	0.009	0.454	0.004	0.434
Days to receipt	0.005	0.584	0.009	0.365	0.004	0.445
Antibiotic use					0.018	0.003

Table 3. Temporal reliability of microbiome measures. Intraclass correlation coefficients for interquartile log-ratio transformed phyla, alpha diversity, and beta diversity measures.

	No antibiotic use (n=27)	Antibiotic use (n=23)	Total (n=50)
Phylum			
Firmicutes	0.64	0.46	0.57
Bacteroidetes	0.62	0.59	0.56
Proteobacteria	0.65	0.44	0.56
Actinobacteria	0.67	0.49	0.59
Alpha diversity			
Phylogenetic diversity	0.75	0.55	0.66
Shannon index	0.67	0.46	0.58
Chao1	0.56	0.45	0.52
Beta diversity			
Unweighted UniFrac PC1	0.93	0.83	0.89
Weighted UniFrac PC1	0.65	0.66	0.64
Bray-Curtis PC1	0.95	0.88	0.90
Jaccard PC1	0.95	0.90	0.90

Figure legends

Figure 1. Variability of the gut microbiome. Principal coordinate plots based on unweighted UniFrac for no antibiotic use (A), antibiotic use (B), all participants (C), and weighted UniFrac for no antibiotic use (D), antibiotic use (E), all participants (F). Smaller dots indicate samples and are connected to larger dots which represent the mean PC1 and PC2 values for each individual. Dashed lines are connected to samples with reported antibiotic use.

Figure 2. Reliability of genera. Intraclass correlations (ICC) genera were calculated for IQLR-transformed abundances for no antibiotic use (A), antibiotic use (B), all participants (C), and presence/absence for no antibiotic use (D), antibiotic use (E), all participants (F) and are plotted against mean abundance. Dotted line indicates ICC of 0.40.

Figure 3. Microbiome recovery from antibiotics. Recovery was assessed among participants reporting use of antibiotics during the two years of collection. Percent change in Shannon index (A) and phylogenetic diversity (B) was calculated relative to the baseline sample and categorized on time since last reported antibiotic use. Changes in alpha diversity for each time interval were assessed using a one-sample t-test for a mean of zero (*$p<0.05$). Distances from baseline were computed for unweighted UniFrac (C) and weighted UniFrac (D) measures. Comparison of six-month to baseline samples among participants not taking antibiotics are also included.

Figure 1

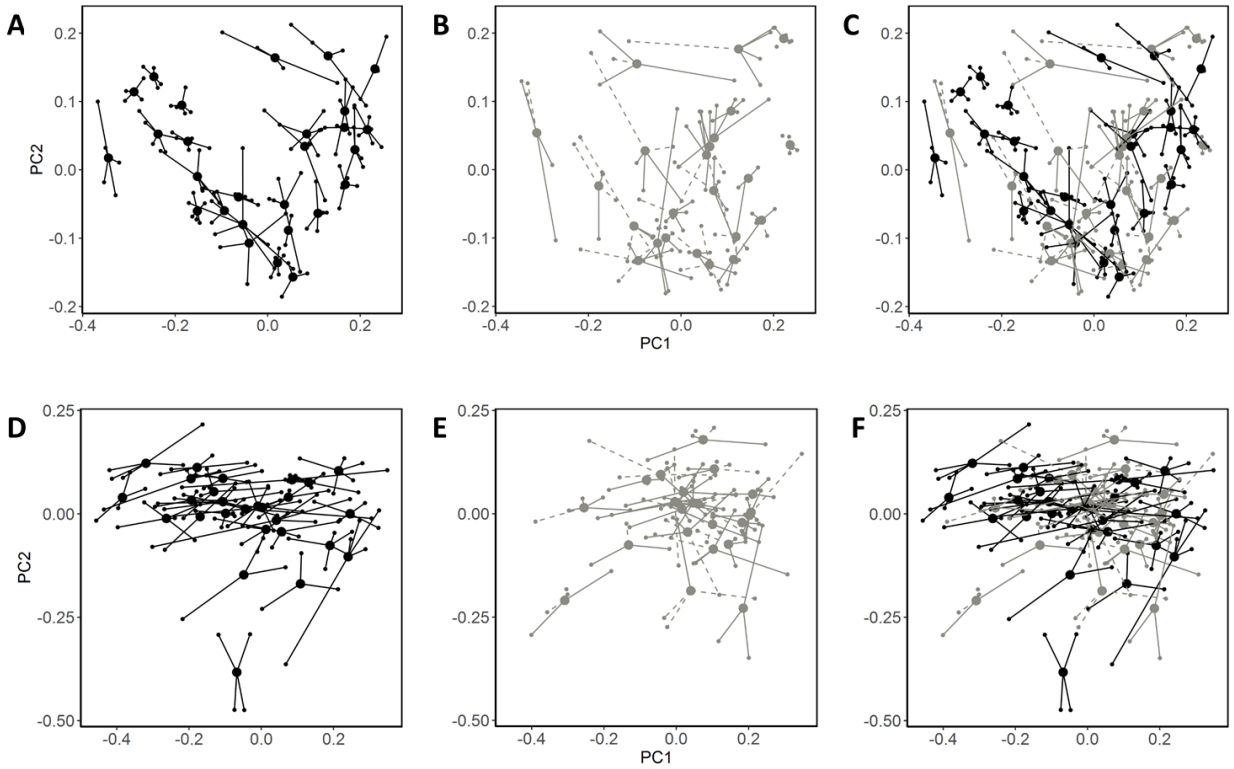


Figure 2

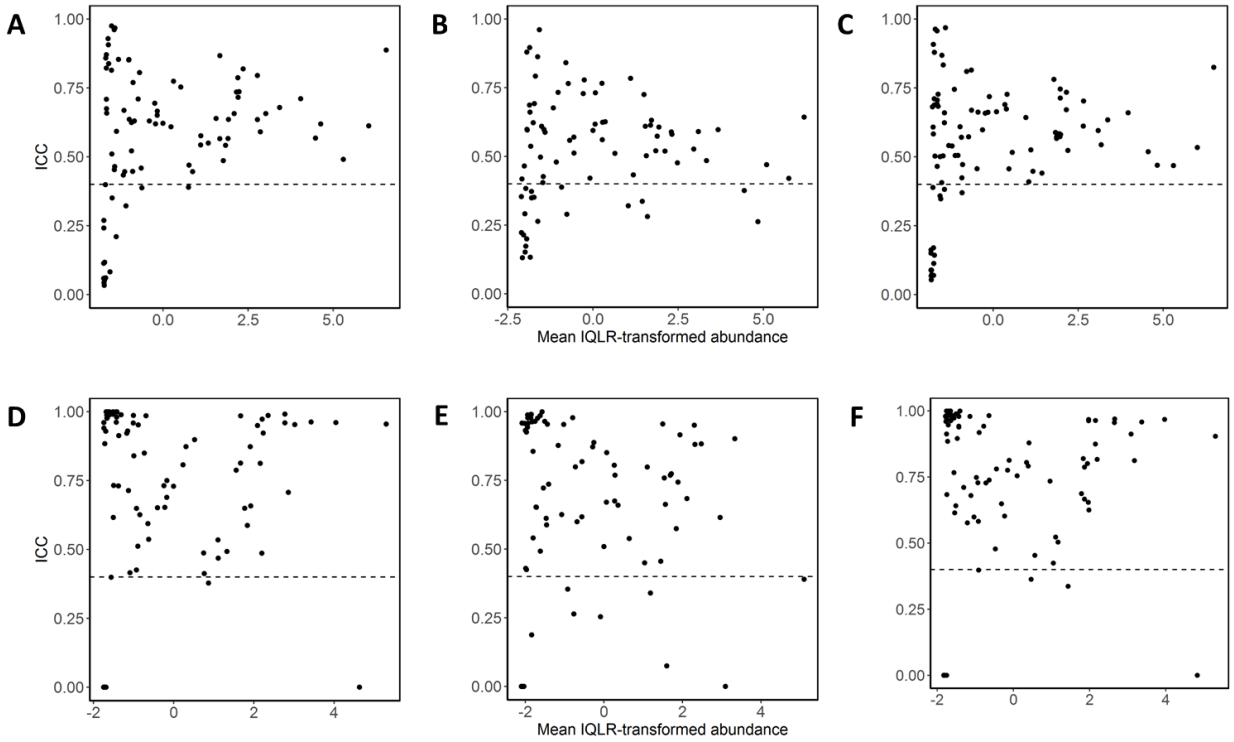
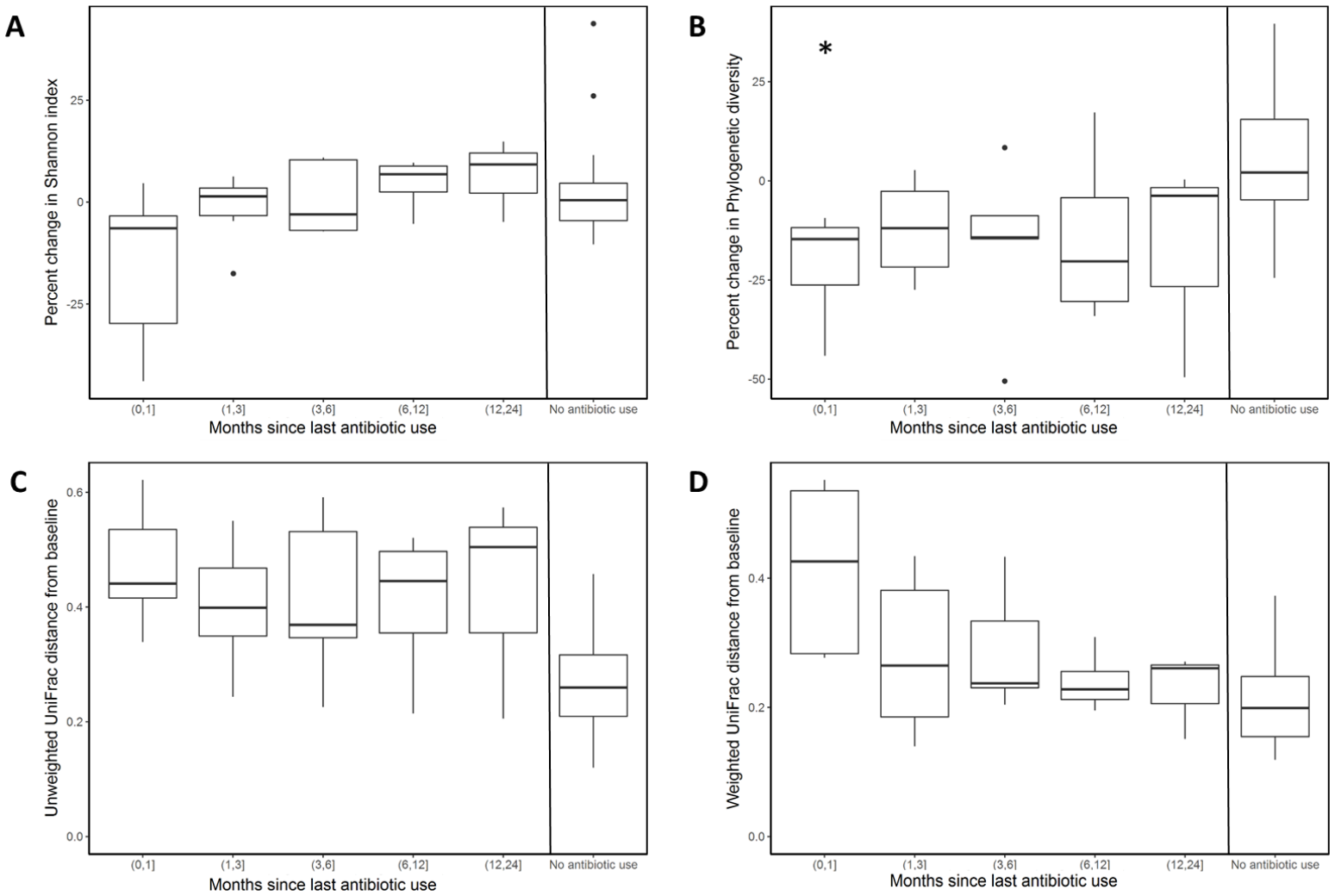

A
B
C
D
E
F
ICC
Mean IQLR-transformed abundance
1.00
0.75
0.50
0.25
0.00
-2.5
0.0
2.5
5.0
-2
0
2
4

Figure 3


A
Percent change in Shannon index
25
0
-25
(0,1]
(1,3]
(3,6]
(6,12]
(12,24]
No antibiotic use
Months since last antibiotic use
B
Percent change in Phylogenetic diversity
*
25
0
-25
-50
(0,1]
(1,3]
(3,6]
(6,12]
(12,24]
No antibiotic use
Months since last antibiotic use
C
Unweighted UniFrac distance from baseline
0.6
0.4
0.2
0.0
(0,1]
(1,3]
(3,6]
(6,12]
(12,24]
No antibiotic use
Months since last antibiotic use
D
Weighted UniFrac distance from baseline
0.4
0.2
0.0
(0,1]
(1,3]
(3,6]
(6,12]
(12,24]
No antibiotic use
Months since last antibiotic use

Supplemental Table 1. Phyla relative abundances and alpha diversity measures for samples collected at baseline, by any reported antibiotic use.

	No antibiotic use (n=27)	Antibiotic use (n=23)	Total (n=50)
Phylum			
Firmicutes	0.55 ± 0.13	0.59 ± 0.14	0.57 ± 0.14
Bacteroidetes	0.38 ± 0.11	0.34 ± 0.13	0.36 ± 0.12
Proteobacteria	0.03 ± 0.08	0.02 ± 0.02	0.02 ± 0.06
Actinobacteria	0.01 ± 0.02	0.01 ± 0.01	0.01 ± 0.01
Alpha diversity			
Phylogenetic diversity	10.27 ± 2.03	11.03 ± 1.28	10.62 ± 1.75
Shannon index	5.72 ± 0.82	5.82 ± 0.53	5.77 ± 0.70
Chao1	187.41 ± 47.88	197.84 ± 33.50	192.23 ± 41.80

Supplemental Table 2. Intraclass correlation coefficients and mean IQLR-transformed abundances for genera among all participants.

Phylum__Class__Order__Family__Genus	Mean IQLR abundance	ICC
Bacteroidetes__Bacteroidia__Bacteroidales__Prevotellaceae__uncultured	-1.40	0.97
Fusobacteria__Fusobacteriales__CFT112H7__uncultured_bacterium__Other	-1.65	0.96
Proteobacteria__Gammaproteobacteria__Aeromonadales__Succinivibrionaceae__Succinivibrio	-1.70	0.96
Bacteroidetes__Bacteroidia__Bacteroidales__S24.7__uncultured_bacterium	-1.76	0.91
Bacteroidetes__Bacteroidia__Bacteroidales__Rikenellaceae__RC9_gut_group	-1.73	0.88
Firmicutes__Erysipelotrichi__Erysipelotrichales__Erysipelotrichaceae__Other	-1.51	0.87
Firmicutes__Erysipelotrichi__Erysipelotrichales__Erysipelotrichaceae__Catenibacterium	-1.47	0.83
Bacteroidetes__Bacteroidia__Bacteroidales__Bacteroidaceae__Bacteroides	6.49	0.82
Bacteroidetes__Bacteroidia__Bacteroidales__Porphyromonadaceae__Barnesiella	-0.64	0.81
Bacteroidetes__Bacteroidia__Bacteroidales__Prevotellaceae__Other	-0.78	0.81
Bacteroidetes__Bacteroidia__Bacteroidales__Prevotellaceae__Prevotella	1.79	0.78
Firmicutes__Clostridia__Clostridiales__Ruminococcaceae__Subdoligranulum	1.97	0.75
Firmicutes__Clostridia__Clostridiales__Veillonellaceae__Dialister	-1.14	0.74
Bacteroidetes__Bacteroidia__Bacteroidales__Rikenellaceae__Alistipes	2.16	0.73
Proteobacteria__Betaproteobacteria__Burkholderiales__Alcaligenaceae__Parasutterella	0.41	0.73
Bacteroidetes__Bacteroidia__Bacteroidales__Porphyromonadaceae__Butyricimonas	-1.62	0.73
Bacteroidetes__Bacteroidia__Bacteroidales__Porphyromonadaceae__Odoribacter	-0.11	0.72
Firmicutes__Clostridia__Clostridiales__Veillonellaceae__Phascolarctobacterium	1.97	0.71
Proteobacteria__Deltaproteobacteria__Desulfovibrionales__Desulfovibrionaceae__Desulfovibrio	-1.64	0.71
Lentisphaerae__Lentisphaeria__Victivallales__Victivallaceae__Victivallis	-1.73	0.71
Firmicutes__Clostridia__Clostridiales__Ruminococcaceae__uncultured	2.66	0.70
Firmicutes__Clostridia__Clostridiales__Christensenellaceae__uncultured	0.35	0.69
Tenericutes__Mollicutes__RF9__Other__Other	-1.65	0.69
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__Anaerosporobacter	-1.73	0.69
Verrucomicrobia__Opitutae__Other__Other__Other	-1.62	0.68

Bacteroidetes__Other__Other__Other__Other	-1.78	0.68
Bacteroidetes__Bacteroidia__Bacteroidales__Porphyromonadaceae__Parabacteroides	2.15	0.67
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__Howardella	0.39	0.67
Firmicutes__Clostridia__Clostridiales__Ruminococcaceae__Oscillibacter	-0.63	0.67
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__Blautia	3.97	0.66
Proteobacteria__Betaproteobacteri__Burkholderiales__Alcaligenaceae__Sutterella	0.10	0.66
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__uncultured_bacterium	-0.15	0.66
Bacteroidetes__Bacteroidia__Bacteroidales__Prevotellaceae__Paraprevotella	-0.23	0.66
Firmicutes__Clostridia__Clostridiales__Ruminococcaceae__Flavonifractor	-0.44	0.66
Cyanobacteria__4C0d.2__uncultured_bacterium__Other__Other	-1.43	0.66
Actinobacteria__Coriobacteriia__Coriobacteriales__Coriobacteriaceae__Collinsella	0.96	0.64
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__Anaerostipes	3.37	0.63
Firmicutes__Clostridia__Clostridiales__Veillonellaceae__Acidaminococcus	-1.43	0.62
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__Dorea	2.65	0.61
Actinobacteria__Bifidobacteriales__Bifidobacteriaceae__Bifidobacterium__uncultured_bacterium	-0.97	0.61
Firmicutes__Clostridia__Clostridiales__uncultured__Other	-1.76	0.61
Firmicutes__Clostridia__Clostridiales__Ruminococcaceae__Other	3.09	0.60
Firmicutes__Clostridia__Clostridiales__Peptostreptococcaceae__uncultured	-0.31	0.60
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__Moryella	1.83	0.59
Firmicutes__Erysipelotrichi__Erysipelotrichales__Erysipelotrichaceae__uncultured	1.98	0.58
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__Coprococcus	1.94	0.58
Firmicutes__Clostridia__Clostridiales__Eubacteriaceae__Other	-1.76	0.58
Bacteria__Other__Other__Other__Other__Other	1.97	0.57
Firmicutes__Clostridia__Clostridiales__Ruminococcaceae__Ruminococcus	1.87	0.57
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__Roseburia	1.87	0.57
Actinobacteria__Bifidobacteriales__Bifidobacteriaceae__Bifidobacterium__human_gut_metagenome	-0.72	0.57
Firmicutes__Clostridia__Clostridiales__Ruminococcaceae__Anaerotruncus	-0.93	0.57
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__Lachnospira	3.18	0.54
Firmicutes__Clostridia__Clostridiales__Ruminococcaceae__Oscillospira	-1.21	0.54

Firmicutes__Clostridia__Clostridiales__Clostridiaceae__Clostridium	-1.30	0.54
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__Other	6.00	0.53
Bacteroidetes__Bacteroidia__Bacteroidales__Other__Other	1.11	0.53
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__Incertae_Sedis	4.56	0.52
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__uncultured	2.20	0.52
Firmicutes__Erysipelotrichi__Erysipelotrichales__Erysipelotrichaceae__Incertae_Sedis	0.56	0.52
Proteobacteria__Gammaproteobacteria__Other__Other__Other	-1.03	0.51
Actinobacteria__Coriobacteriia__Coriobacteriales__Coriobacteriaceae__uncultured	-1.11	0.50
Proteobacteria__Gammaprotebacteria__oB38__uncultured_bacterium__Other	-1.47	0.50
Firmicutes__Clostridia__Clostridiales__Veillonellaceae__Megasphaera	-1.56	0.50
Firmicutes__Clostridia__Clostridiales__uncultured__uncultured_bacterium	-1.71	0.50
Firmicutes__Clostridia__Clostridiales__Ruminococcaceae__Faecalibacterium	5.30	0.47
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__Pseudobutyrivibrio	4.83	0.47
Proteobacteria__Alphaproteobacteria__Rhodospirillales__Rhodospirillaceae__Thalassospira	-0.90	0.47
Firmicutes__Clostridia__Clostridiales__Veillonellaceae__Megamonas	-1.64	0.47
Firmicutes__Bacilli__Lactobacillales__Streptococcaceae__Streptococcus	0.47	0.46
Actinobacteria__Bifidobacteriales__Bifidobacteriaceae__Bifidobacterium__Other	-0.47	0.46
Firmicutes__Clostridia__Clostridiales__Other__Other	1.17	0.45
Firmicutes__Clostridia__Clostridiales__Ruminococcaceae__Incertae_Sedis	1.44	0.44
Firmicutes__Clostridia__Clostridiales__Ruminococcaceae__Pseudoflavonifractor	-0.92	0.42
Verrucomicrobia__Verrucomicrobiae__Verrucomicrobiales__Verrucomicrobiaceae__Akkermansia	1.05	0.41
Actinobacteria__Coriobacteriia__Coriobacteriales__Coriobacteriaceae__Adlercreutzia	-1.51	0.41
Tenericutes__Mollicutes__RF9__uncultured_bacterium_adhufec202__Other	-1.77	0.39
Tenericutes__Mollicutes__RF9__uncultured_bacterium__Other	-1.43	0.38
Firmicutes__Clostridia__Clostridiales__Veillonellaceae__Other	-0.91	0.37
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__Marvinbryantia	-1.56	0.36
Firmicutes__Clostridia__Clostridiales__Veillonellaceae__Veillonella	-1.54	0.35
Proteobacteria__Gammaproteobacteria__Enterobacteriales__Enterobacteriaceae__Enterobacter	-1.75	0.17
Proteobacteria__Alphaproteobacteria__Rhizobiales__Methylobacteriaceae__Methylobacterium	-1.83	0.16

Bacteroidetes__Bacteroidia__Bacteroidales__ratAN060301C__uncultured_bacterium	-1.83	0.15
Proteobacteria__Gammaproteobacteria__Enterobacteriales__Enterobacteriaceae__Escherichia.Shigella	-1.73	0.14
Firmicutes__Clostridia__Clostridiales__Family_XIII_Incertae_Sedis__Other	-1.75	0.11
Actinobacteria__Propionibacteriales__Propionibacteriaceae__Propionibacterium__uncultured_actinobacterium	-1.81	0.09
Proteobacteria__Betaproteobacteria__Burkholderiales__Burkholderiaceae__Burkholderia	-1.82	0.09
Proteobacteria__Gammaproteobacteria__Pseudomonadales__Moraxellaceae__Moraxella	-1.82	0.09
Proteobacteria__Gammaproteobacteria__Pseudomonadales__Pseudomonadaceae__Pseudomonas	-1.75	0.07
Actinobacteria__Micrococcales__Micrococcaceae__Micrococcus__Other	-1.80	0.07
Firmicutes__Bacilli__Bacillales__Staphylococcaceae__Staphylococcus	-1.82	0.07
Proteobacteria__Alphaproteobacteria__Rhizobiales__Other__Other	-1.82	0.05

Supplemental Table 3. Gut microbiome-body fat association analysis calculated using PERMANOVA with unweighted and weighted UniFrac distance metrics. The variation in gut microbiome composition associated with percent body fat was calculated at baseline, followed by the addition of subsequent samples.

	Stool Collection Time Points				
	Baseline	Baseline+6mo	Baseline+6mo+ 12mo	Baseline+6mo+ 12mo+18mo	All samples
No antibiotic use					
Unweighted R^2	0.056	0.043	0.038	0.035	0.035
Weighted R^2	0.045	0.034	0.026	0.028	0.026
Antibiotic use					
Unweighted R^2	0.070	0.066	0.060	0.057	0.056
Weighted R^2	0.127	0.101	0.090	0.088	0.092
All participants					
Unweighted R^2	0.032	0.034	0.031	0.030	0.029
Weighted R^2	0.053	0.037	0.027	0.024	0.025

CHAPTER 3

Associations of plasma trimethylamine *N*-oxide and precursors with inflammatory and cardiometabolic risk biomarkers and the fecal microbiome in the Multiethnic Cohort

Abstract

Background: Trimethylamine N-oxide (TMAO), a compound derived from diet and metabolism by the gut microbiome, has been associated with several chronic diseases, although its disease mechanisms and the bacteria involved in its production are not fully known.

Methods: We assessed relationships between TMAO and its precursors (choline, carnitine, and betaine) with inflammatory and cardiometabolic risk biomarkers and the fecal microbiome within the Multiethnic Cohort. Plasma concentrations of TMAO and its precursors were measured with LC-MS/MS. Biomarkers included inflammatory markers, insulin resistance, HDL and LDL cholesterol, triglycerides, and systolic and diastolic blood pressure. Composition of the fecal microbiome was evaluated by sequencing the 16S rRNA gene V1-V3 region. Associations between TMAO and its precursors with disease biomarkers were assessed by multivariable regression, while associations between TMAO and the fecal microbiome were assessed by PERMANOVA and LASSO regression.

Results: Choline was associated with an adverse risk profile while TMAO and its precursors were all associated with insulin resistance. TMAO explained a small percentage of overall fecal microbiome variation (<1%), but was associated with several genera, including *Desulfovibrio* and ones within the family Lachnospiraceae.

Conclusions: Plasma TMAO concentrations were associated with choline-metabolizing bacteria, and along with its precursors, may contribute to inflammatory and cardiometabolic risk pathways.

Introduction

There is an increasing recognition of the role of gut microbiome-derived metabolites in disease etiology. The compound trimethylamine *N*-oxide (TMAO) has been found to be associated with several chronic diseases, including cardiovascular disease (CVD) (25), colorectal cancer (29), diabetes (26, 31, 105), and chronic kidney disease (35, 106, 107). Although earlier work has shown involvement by TMAO in atherogenic processes (47), the range of different diseases suggests other mechanisms may be at play. Growing experimental evidence in animal models demonstrates a contribution of TMAO to inflammation (41) and metabolic dysfunction (39), highlighting the need for additional epidemiological research in this area.

TMAO is obtained from diet, both directly from foods such as fish and shellfish (108), as well as through the microbial metabolism of choline, carnitine, and betaine to trimethylamine (TMA) (27, 28, 53), which is oxidized to TMAO by hepatic flavin-containing monooxygenases (FMO), particularly FMO3 (109). Several bacterial species involved in TMA production have been identified by culture-based methods, and include those belonging to the phyla *Firmicutes, Proteobacteria,* and *Actinobacteria* (50, 51, 53, 110). Newer bioinformatic approaches have been utilized to identify bacterial genes involved in these conversions, including the *cutC/D* and *cntA/B* gene clusters which metabolize choline and carnitine, respectively (54, 55). The ability to reduce TMA by other microbes, namely Archaea (111), adds yet another layer of complexity in determining circulating TMAO concentrations. While many of these previous studies indicate which bacteria have the potential to produce TMA in the gut environment, it is less clear whether these will be also associated with TMAO in population-based studies.

In this study, we used data from the Multiethnic Cohort to address two aims: 1) investigate associations between TMAO and its precursors (choline, carnitine, and betaine) with inflammatory and cardiometabolic biomarkers to better understand potential disease

mechanisms of these compounds, and 2) identify associations between TMAO and the fecal microbiome.

Methods

Study participants

The Multiethnic Cohort (MEC) is an ongoing prospective cohort study that recruited 215,251 men and women from Hawaii and Los Angeles, California between 1993 and 1996 (71). MEC participants were aged 45 to 75 at the time of recruitment and were predominantly from five racial-ethnic groups: African American, Japanese American, Latino, Native Hawaiian, and white. The current study includes participants from the Adiposity Phenotype Study (APS), a sub-study that recruited 1,861 from the MEC to investigate associations between multi-omics data and body fat distribution. APS participants were selected as to have a similar distribution of men and women across the five racial-ethnic groups, with stratified sampling based on BMI categories (18.5-21.9, 22.0-24.9, 25.0-26.9, 27.0-29.9, 35.0-40.0 kg/m^2). Additional inclusion and exclusion criteria for the APS have been reported in detail previously (72). Notably, individuals were excluded for current or recent (<2 years) smoking, insulin or thyroid medication, dialysis, serious health conditions, and antibiotics use in the last 3 months, as well as colonoscopy, chemotherapy, or radiation therapy in the last 6 months.

In the current analysis, participants were excluded if they did not have blood samples measured for TMAO or biomarkers (n=9). Participants were also excluded (n=136) if they did not have information for key variables (aspirin use, diet, percent body fat, and physical activity), leaving 1,716 available for the current analysis. Carnitine data were available for 1,434 of these participants.

Questionnaires

Prior to the clinic visit, participants filled out a mailed questionnaire containing items related to demographics, health and medication history, physical activity, and a quantitative food frequency questionnaire (FFQ).(112) Metabolic Equivalent of Tasks (METs) were calculated based on reported average time spent in light, moderate, and strenuous activity during the past year. Questions related to usual eating habits of over 180 food items during the last year were included in the FFQ, which has been validated and calibrated against 24-hour dietary recalls within the MEC, and incorporates many ethnic-specific foods (112). Participants were asked how often they ate each food item (8 possible responses ranging from “never or hardly ever” to “2 or more times a day”) and the usual serving size, which was accompanied by pictures of three different portion sizes to assist in estimation. Food groups were calculated as grams per day based on relevant food items, as well as portions of mixed dishes. The questionnaire was filled out by the participant at home and reviewed by study staff during the clinic visit. For the current analysis, we included food groups and nutrients that are major sources of TMAO (fish and shellfish) and its precursors (red and processed meats, eggs, fiber).

Study clinic visit

At the study clinic visit, participants had anthropometrics measured, fasting blood samples drawn, and underwent dual-energy X-ray absorptiometry (DXA) and magnetic resonance imaging (MRI) scanning. Venous blood was collected after an overnight fast (>8 hours) in two 10 mL heparinized Vacutainer tubes and two 10 mL dry tubes. Fasting blood was processed into components within 4 hours of collection and frozen at -80˚C until analysis at the University of Hawaii Cancer Center (UHCC) Analytical Biochemistry Shared Resource lab. The DXA scan (Hologic Discovery A) was performed to measure total and regional body fat mass (113). DXA image files from both study sites were centrally analyzed at the University of California, San Francisco to estimate percent body fat.

Systolic (SBP) and diastolic blood pressure (DBP) in the left arm was measured in a sitting position after 20 minutes of rest. Trained technicians measured blood pressure in the left arm of the participant using a digital monitor (Omron HEM-907XL, Lake Forest, IL). If the first two measurements differed by 10 mmHg or more, a third measurement was taken and the closest two were averaged.

Blood biomarker analysis

Serum was analyzed on a Cobas MiraPlus chemistry analyzer (Roche, Indianapolis, IN) for glucose using kits from Randox Laboratories (#GL1611) and high-sensitivity C-reactive protein (hsCRP), triglycerides (TG) and total high-density lipoprotein cholesterol (HDL-C) were measured using kits from Pointe Scientific (C7568, T7532, and H7545). Low-density lipoprotein cholesterol (LDL-C) was derived from total cholesterol and HDL-C among individuals with TG below 400 mg/dL using the Friedewald equation (114). Enzyme-linked immunosorbent assays (ELISA) were used to measure serum insulin (EMD Millipore EZHI-14K) and plasma lipopolysaccharide binding protein (LBP; Cell Sciences CKH113). The homeostasis model assessment estimate for insulin resistance (HOMA-IR) was derived from glucose and insulin measurements (115, 116). Blind duplicate QC samples (10% of study samples) were included at random in each batch and yielded coefficients of variation and intra-class correlation coefficients (CVs and ICCs) as follows: glucose: 10% and 90%; hsCRP: 55%, 86%; TG: 18%, 95%; total cholesterol: 14%, 72%; HDL-C: 18%, 80%; insulin: 14%, 96%; and LBP: 18%, 76%.

Plasma choline, betaine, carnitine, and TMAO analysis

Plasma choline, betaine, carnitine, and TMAO were analyzed by tandem LCMS (Surveyor HPLC coupled to TSQ Quantum™ mass spectrometer, Thermo Scientific Inc., Waltham, MA) after electrospray ionization in positive mode using selected reaction monitoring modified from published methods (117, 118). Plasma (0.025 mL) was mixed with 0.01 mL of an

aqueous internal standard solution consisting of choline-(trimethyl-d_9) hydrochloride, betaine-(trimethyl-d_9) hydrochloride, L-carnitine-d_3 hydrochloride (all from Sigma, St. Louis, MO) and trimethylamine-d_9 N-oxide, (Cambridge Isotopes, Tewksbury, MA). Proteins were then precipitated by the addition of 0.215 mL acetonitrile. After vortexing for 5 min. and centrifuging for 5 min. at 14,000 x g, 0.01 mL of the clear supernatant was injected onto a ZIC-c HILIC column (100 x 2.1mm, 3µm, 100A) with a ZIC-c HILIC guard column (20 x 2.1mm; The Nest Group, Southborough, MA). The mobile phases consisted of A=MeCN/EtOH/H_2O/100mM NH_4OAc, pH 4.8 (40/6.8/40/13.2; v/v/v/v) and B= MeCN/EtOH/H_2O/100mM NH_4OAc, pH 4.8 (80/6.8/12.7/0.5; v/v/v/v). Linear gradient elution was performed at a flow rate of 0.2 mL /min. as follows (%A): 0-2.0 min. at 20%, 2.0-5.0 min. to 100%, 5.0-6.0 min. hold at 100%, 6.0-6.1 min. linear gradient to 20% and equilibrate at 20% for 5.9 minutes.

The general MS conditions were as follows: source, ESI; ion polarity, positive; spray voltage, 4500 V; sheath and auxiliary and Ion sweep gas, nitrogen; sheath gas pressure, 45 arbitrary units; auxiliary gas pressure, 5 arbitrary units; ion sweep gas pressure, 0 arbitrary units; ion transfer capillary temperature, 350°C; scan type, high resolution selected reaction monitoring; collision gas, argon; collision gas pressure, 1.0 mTorr, source CID 5V; scan width, 0.01u; scan time, 1 s; Q1 peak width was set at 0.7 u full width at half maximum (FWHM) and Q3 peak width at 0.70 u FWHM. Mass spectrometric monitoring is started 0.0 minutes after sample injection by multiple reaction monitoring using transitions (only ions quantitated on are listed, collision energies applied in brackets) for choline from *m/z* 104.109 to *m/z* 45.203 (21eV), 58.186 (33eV), and 60.188 (16eV); for betaine from *m/z* 118.089 *to m/z* 42.216 (53eV), 58.181 (25eV), and 59.186 (18eV); for choline trimethyl-d_9 from *m/z* 113.000 to *m/z* 49.200 (35eV), 66.203 (32eV), and 69.235 (17eV); for betaine trimethyl-d_9 from *m/z* 127.000 to *m/z* 64.217 (35eV), 66.219 (30eV), and 68.209 (20eV) for trimethylamine N-oxide from *m/z* 76 to *m/z* 42.276 (42eV), 58.205 (19eV) for trimethylamine N-oxide-d_9 from *m/z* 85.00 *to m/z* 46.284 (38eV), 66.228 (19eV), and 68.276 (13eV); for carnitine from *m/z* 162.098 to *m/z* 60.176 (16eV),

85.086 (20eV), and 103.083 (16eV); and for carnitine-d_3 from *m/z* 165.060 to *m/z* 61.209 (40eV), 103.128 (16eV), and 105.154 (18eV). Final concentrations were obtained by external calibration.

Microbiome sample collection and processing

Participants received a stool collection kit during their clinic visit. Stool samples were collected at the participants' home into a vial containing RNAlater and frozen overnight (73). Participants were then asked to bring the sample to the UHCC or University of Southern California (USC) study center. The UHCC and USC labs stored the samples at -80°C until bulk shipments were made every 3 months to the Fred Hutchinson Cancer Research Center (Fred Hutch), where they were stored at -80°C until processing.

Laboratory and bioinformatic processing procedures have been previously described (please refer to the Methods section in Chapter 2). Briefly, DNA from stool samples was extracted at Fred Hutch, amplified for the V1-V3 region of the 16S rRNA gene, and shipped to Research and Testing Laboratory (RTL), LLC (Lubbock, TX) for sequencing. Gut microbial composition of stool samples was assessed with 2x300 bp paired-end sequencing on the Illumina MiSeq platform. Quality control of sequences and inference of phylogenetic relationships were done using Quantitative Insights Into Microbial Ecology (QIIME) v1.8 (74) pipelines. All failed sequence reads and low-quality sequence ends were filtered. Chimeric and non-bacterial sequences were also removed. Filtered sequences were grouped into Operational Taxonomic Units (OTUs) at 97% similarity and aligned for phylogenetic analysis. Alpha and beta diversity measures were rarefied to 9,000 sequences per sample.

Real-time PCR quantification of Methanobrevibacter smithii

Total bacteria was quantified using TaqMan real-time PCR (Applied Biosystems, QuantStudio 5) 8FM (5'-AGAGTTTGATCMTGGCTCAG-3')-530R primers (5'-

TTACCGCGGCKGCTGGCAC-3') and a Bac 338 NED labeled probe (5'-CCAKACTCCTACGGGAGGCAGCAG-3') (Applied Biosystems). For quality control purposes, we used a standard made of DNA extracted from the following pure cultures: *B. fragilis* HM-20D BEI Resources, *B. adolescentis* 15703D ATCC, *C. difficile* BAA-1382D-5 ATCC, *V. parvula* 10790D-5 ATCC, and *E. coli*. The TaqMan PCR assays were carried out in a total reaction volume of 10 µL using 5.0 µL of 2X TaqMan Multiplex Master Mix (Applied Biosystems) containing 2µL of templet DNA, 0.025 µM of each probe, and 0.125 µM of each primer. The PCR thermal-cycling conditions of the DNA were set at 50 °C for 2 min, 95 °C for 10 min, and 40 cycles of 95 °C for 15 s and 60 °C for 1 min (119). The archaea *Methanobrevibacter smithii* was measured using PCR primers directed at archaeal 16S rRNA genes [FW 5′-CCGGGTATCTAATCCGGTTC-3′; and 5′-CTCCCAGGGTAGAGGTGAAA-3′; and FAM labeled probe F FAM: 5′-CCGTCAGAATCGTTCCAGTCAG-3′] with the following cycling conditions: 30 cycles of denaturing at 94 °C for 2 min, annealing at 65 °C for 45 s, and extension at 72 °C for 2 min. DNA extracted from a pure culture was used to generate standard curves with the same primer probe set (*Methanobrevibacter smithii*, DSM 11975). Data were analyzed using QuantStudio software (v1.2.x; Applied Biosystems). *M. smithii* was expressed as a ratio of total bacteria in the sample (120).

Statistical analysis

Descriptive statistics of the study population were calculated for each quartile of plasma TMAO concentration. Associations of plasma TMAO, choline, carnitine, and betaine with LBP, HOMA-IR, HDL cholesterol, LDL cholesterol, triglycerides, SBP, and DBP were determined by multivariable linear regression. For CRP, we applied a zero-inflated negative binomial (ZINB) regression model using the 'zeroinfl' function in the R package pscl (121) due to the number of participants with CRP values of 0. Both linear and ZINB regression models were adjusted for the following confounding factors selected *a priori*: age (continuous), sex, race-

ethnicity (five groups), physical activity (metabolic equivalents; METs), percent body fat (continuous), and aspirin use (no, previous, current). TMAO, choline, betaine, and carnitine, as well as HOMA-IR and triglyceride concentrations, were log-transformed for all models.

To identify associations between plasma TMAO and the fecal microbiome, we examined microbiome community structure as a whole, as well as at genus-level taxa. Both analyses included adjustment of laboratory batch. The variation in the microbiome explained by TMAO and its precursors was assessed by PERMANOVA (999 permutations) for both unweighted and weighted UniFrac distance measures with the 'adonis2' functions in the R package 'vegan' (122). Marginal effects were calculated with laboratory batch included as a covariate in the model using the "margin" option. We used least absolute shrinkage and selection operator (LASSO) regression (123) as a method of variable selection to identify genera associated with TMAO. Genera were centered log-ratio transformed and standardized, and only those present in at least 20% of participants were included in the model. Laboratory batch was included in the model by setting the penalty factor argument to 0. The LASSO tuning parameter was chosen by 10-fold cross-validation.

We also used linear regression, adjusting for age, sex, race-ethnicity, and percent body fat, to assess the relationship of TMAO and choline separately with *M. smithii*. *M. smithii* was evaluated as a continuous variable and as a binary variable (presence/absence).

Statistical tests were two-sided with significance at $p<0.05$. Adjusted p-values for the biomarker analysis were also calculated to correct for the false discovery rate (FDR) by using the Benjamini-Hochberg procedure (124). Analyses were conducted in R version 3.4.4.

Results

Participant characteristics

Among the 1716 study participants, those with higher plasma TMAO were more likely to be current users of aspirin and have lower physical activity (Table 1). Higher concentrations of

choline and carnitine were seen in the upper quartiles of TMAO, with a similar pattern across race-ethnic groups (Supplemental Table 1). Dietary intake of fish, shellfish, and processed meats, but not eggs, also differed across TMAO quartiles. Fish and shellfish exhibited a nonlinear trend, with the lowest intake in the second quartile. Processed meat showed an increasing trend, with the highest intake in the fourth quartile.

Associations between TMAO compounds and biomarkers

We identified several associations between TMAO, choline, carnitine, and betaine with inflammatory and cardiometabolic biomarkers (Table 2). Plasma choline (β=0.23, p=0.03) and betaine (β=0.29, p=0.02), but not TMAO or carnitine, were positively associated with CRP, whereas only choline was associated with LBP (β=1968.3, p=0.01). All four plasma compounds were associated with HOMA-IR, with only betaine having an inverse association (β=-0.32, p<0.001). Choline was inversely associated with HDL cholesterol (β=-14.86, p<0.001) and positively associated with LDL cholesterol (β=9.86, p=0.049). Choline (β=0.14, p=0.002), carnitine (β=0.18, p<0.001), and betaine (β=-0.16, p<0.001) were associated with triglycerides. Betaine was inversely associated with systolic (β=-2.68, p=0.04) and diastolic blood pressure (β=-2.27, p=0.003), while choline was positively associated with systolic blood pressure (β=4.50, p=0.01). A majority of significant associations remained significant after correcting for multiple comparisons. Parameter estimates were similar after further adjustment for energy intake and dietary intake of seafood and red and processed meat (Supplemental Table 2).

TMAO and the fecal microbiome

PERMANOVA analysis showed that TMAO, choline, carnitine, and betaine each explained between 0.1-0.3% of the variation in the fecal microbiome (Table 3). TMAO (R^2=0.0010, p=0.007), choline (R^2=0.0010, p=0.0006), and carnitine (R^2=0.0018, p=0.002) were

significantly associated with the microbiome for unweighted UniFrac, and TMAO (R^2=0.0014, p=0.038) and carnitine (R^2=0.0029, p=0.007) was associated for weighted UniFrac.

Out of 104 genera present in the sample of APS participants, 92 were present in at least 20% of participants. LASSO regression with these genera identified 4 which were positively associated with TMAO (*Clostridia*, *Blautia*, an uncultured bacterium within *Peptostreptococcaceae*, and *Desulfovibrio*) and 6 which were inversely associated (*Barnesiella*, an uncultured *Peptostreptococcaceae* genus, *Faecalibacterium*, an uncultured *Clostridiales* genus, *Halocella*, and *Clostridium* belonging to the order Erysipelotrichales) (Table 4). Two genera positively associated with TMAO, belonging to the family Lachnospiraceae and Peptostreptococcaceae, were also positively associated with choline (Supplemental Table 3).

M. smithii was not associated with TMAO when modeled as a continuous variable (Supplemental Table 4), although there was a positive association for presence of *M. smithii* (β=0.071, p=0.043). There was no association between *M. smithii* and choline.

Discussion

In the present study, we identified several associations between TMAO and its precursors choline, carnitine, and betaine, with inflammatory and cardiometabolic biomarkers. Notably, we found that choline was significantly associated with nearly all markers and indicated risk to adverse health outcomes, and that TMAO and its precursors were all associated with insulin resistance. We also showed that TMAO and its precursors explained only a small proportion of fecal microbiome variation (<1%), although several genera were associated with TMAO concentration.

While the exact functions of TMAO in the human body have yet to be elucidated, its precursors are known to play critical roles in health and function. Carnitine is involved in the oxidation of fatty acids (125), while choline is an essential nutrient necessary for production of the neurotransmitter acetylcholine, as well as for phospholipids in cell membranes (126).

Betaine, which can be obtained directly from diet as well as by metabolism of choline, serves as a methyl donor for the conversion of homocysteine to methionine (127). Despite having critical roles in metabolic pathways, there is epidemiological evidence that these compounds may also contribute to cardiometabolic disease risk. Our results for plasma choline and betaine in particular are comparable to previous studies, which have found unfavorable cardiovascular risk profiles for higher concentrations of choline and lower concentrations of betaine. Roe et al. recently reported levels of plasma choline to be associated with higher homocysteine concentrations and body mass index and lower HDL cholesterol, while betaine was associated with lower LDL cholesterol and triglycerides, as well as lower odds of diabetes (45). Similar patterns have been seen in studies conducted in Norway (44) and New Zealand (46).

These consistent associations across studies with components of metabolic syndrome, along with insulin resistance in our study, suggest a potential role of these compounds in diabetes. In fact, both dietary phosphatidylcholine intake and plasma TMAO have been linked to significantly increased type 2 diabetes risk in population-based studies (31, 128). There has also been support in animal models; Gao and colleagues showed that dietary supplementation of mice with TMAO led to impaired glucose tolerance, alterations in hepatic insulin signaling pathways, and promotion of adipose tissue inflammation (39). Additionally, insulin suppressed expression of FMO3 *in vitro*, and knockdown of the enzyme in insulin-resistant mice suppressed FOXO1, a key transcription factor involved in the regulation of insulin signaling (43).

Interestingly, we did not find TMAO to be associated with any other disease biomarker, even though a variety of studies have linked TMAO with CVD and CVD-related mortality (47). It should be noted, however, that many of these studies have been conducted in participants with other diseases or patient populations undergoing procedures such as hemodialysis and cardiac catheterization. Thus, these associations may in part be influenced by other health outcomes or confounding factors. Although our participants were older age, they were relatively healthy with no serious health conditions and were not on dialysis or undergoing insulin treatment. Additional

studies in healthier populations are needed and may also show divergent results, as was seen in a prospective cohort study of healthy adults which did not find associations between TMAO and measures of atherosclerosis (129).

How TMAO is obtained (i.e. directly from food or derived from precursors) could also impact these associations. For example, fish is a major source of TMAO even though fish consumption has been shown to reduce risk of CVD (130). The potential effects of TMAO on cardiometabolic pathways may be counteracted by other compounds present in fish. The dietary intake of fish in our study is likely greater than in previous studies that have been primarily in white and black populations, given the higher consumption among Japanese-Americans and Native Hawaiians (71, 131). We took into account food sources of TMAO by further adjusting for these in our model and found similar associations, although this may not have fully captured the circulating levels of specific compounds that impact disease risk.

As gut microbes are necessary for the conversion of choline, carnitine, and betaine into TMA (28, 53), we hypothesized that composition of the gut microbiome would be associated with TMAO concentrations. We found that TMAO and its precursors explained <1% of overall fecal microbiome variation, which may not be surprising given that TMA-producing bacteria represent a small component of the gut bacterial community (56). Furthermore, the association may be influenced by variation of other factors related to the human conversion of TMA to TMAO as well as excretion rates of TMAO. Carnitine actually had the strongest association among the four compounds, possibly reflecting meat or protein intake, or overall dietary patterns (65, 132).

We found several genera associated with TMAO using LASSO regression, including *Desulfovibrio*, from which the *cut* gene cluster was initially discovered.(54) Two genera belonging to the family *Lachnospiraceae* were also positively associated with TMAO in our study. Several bacteria in *Lachnospiraceae* have been identified to possess *cutC/D*, especially those in *Clostridium XIVa* (56), *Lachnospiraceae* was also more abundant among high-TMAO

producers in a crossover feeding trial of healthy young men (108). None of the genera we identified are known to have *cntA/B*, which may reflect the fact that the gene cluster is less prevalent in gut bacteria (53, 56).

Archaea also play a role in determining TMAO concentrations. Using TMA, members of the Methanomassiliicoccales order can produce methane through the trimethylamine methyltransferase (*mttB)* gene (111, 133). This has been shown experimentally, as the strain *Methanomassiliicoccus luminyensis B10* was able to reduce TMA and H_2 for methanogenesis (111). Although *M. smithii* does not possess the *mttB* gene and thus would not directly impact TMA or TMAO concentrations through this pathway, we found that participants with the Archaea had increased levels of TMAO. Research in the role of Archaea in disease risk is still in its infancy, and additional studies are needed to better understand the role of *M. smithii* and other Archaea in contributing to production of TMAO.

Our study has several strengths and limitations. First, we had an ethnically diverse and relatively healthy sample, whereas many of the TMAO studies to date have been limited to white and/or black populations of participants with disease or undergoing various medical procedures. Second, we were also able to assess associations between disease biomarkers with TMAO, as well as choline, carnitine, and betaine. Third, this is one of few population-based studies of TMAO and the gut microbiome, and is the largest to-date. A limitation of our study was the cross-sectional design, so we were unable to infer causality between TMAO and its precursors with disease risk biomarkers. We also used 16S rRNA gene data, which did not allow us to examine associations between TMAO and bacterial functional genes, such as the *cut* and *cnt* clusters.

In summary, our findings lend support to a possible role of TMAO and its precursors in cardiometabolic risk, as well as associations between TMAO and fecal bacteria. Longitudinal studies of TMAO and its precursors in disease risk are needed, along with additional population-based studies of TMAO and the gut microbiome.

Table 1. Characteristics of study participants by quartiles of TMAO

	TMAO Quartile			
	Q1 <2.10 µmol/L	Q2 2.10-3.05 µmol/L	Q3 3.06-4.60 µmol/L	Q4 >4.60 µmol/L
N	432	427	431	426
Age (years)	69.2 ± 2.8	69.0 ± 2.7	69.2 ± 2.8	69.3 ± 2.7
Female (%)	228 (52.8)	220 (51.5)	208 (48.3)	200 (46.9)
Race-ethnicity (%)				
African American	72 (16.7)	75 (17.6)	63 (14.6)	65 (15.3)
Native Hawaiian	71 (16.4)	69 (16.2)	68 (15.8)	73 (17.1)
Japanese American	126 (29.2)	91 (21.3)	94 (21.8)	118 (27.7)
Latino	96 (22.2)	85 (19.9)	82 (19.0)	72 (16.9)
White	67 (15.5)	107 (25.0)	124 (28.8)	98 (23.0)
Cigarette smoking history (%)				
Former	268 (62.0)	259 (60.7)	258 (59.9)	256 (60.1)
Never	164 (38.0)	168 (39.3)	173 (40.1)	170 (39.9)
Physical activity (METs)	1.7 ± 0.3	1.7 ± 0.3	1.6 ± 0.3	1.6 ± 0.3
Body fat (%)	33.4 ± 7.7	33.4 ± 7.9	33.2 ± 7.6	33.3 ± 7.8
Aspirin use				
No	264 (61.1)	234 (54.8)	228 (52.9)	204 (47.9)
Previous	55 (12.7)	54 (12.6)	68 (15.8)	61 (14.3)
Current	113 (26.2)	139 (32.6)	135 (31.3)	161 (37.8)
TMAO precursors (µmol/L)				
Choline	11.5 [9.9-13.3]	12.3 [10.6-14.2]	12.8 [10.8-14.9]	13.1 [11.0-15.8]
Betaine	39.9 [31.7-47.7]	40.3 [33.5-49.2]	41.2 [33.6-51.1]	40.4 [31.4-48.8]
Carnitine	38.7 [32.7-44.3]	38.7 [32.9-44.0]	39.2 [34.2-45.5]	40.9 [35.6-46.8]
Total energy (kcal/day)	1675 [1303-2230]	1690 [1221-2170]	1753 [1334-2323]	1719 [1272-2291]
Dietary intake (g/day)				
Fish	16.7 [7.0-28.6]	15.5 [7.2-27.2]	17.5 [7.5-32.7]	18.8 [10.1-33.4]
Shellfish	2.0 [0.7-4.1]	1.6 [0.5-3.6]	2.0 [0.6-4.3]	2.1 [0.8-4.9]
Fish + shellfish	19.0 [8.7-32.0]	17.8 [8.6-30.8]	20.1 [9.0-36.9]	21.5 [12.2-39.0]
Red meat	22.6 [12.2-39.6]	20.7 [12.4-36.6]	22.3 [11.8-38.7]	23.4 [12.8-39.7]
Processed meat	11.5 [5.2-21.2]	11.4 [5.5-19.0]	12.1 [5.4-23.3]	13.5 [7.0-24.2]
Red + processed meat	35.6 [19.1-64.3]	34.3 [19.1-56.3]	35.3 [21.3-60.4]	39.3 [22.4-65.9]
Eggs	16.4 [9.6-30.2]	15.5 [8.3-27.8]	17.7 [8.7-34.5]	15.8 [9.2-26.6]
Fiber	19.8 [13.6-28.9]	20.4 [14.1-28.0]	21.3 [14.8-29.3]	19.6 [14.0-28.6]

Values presented as n(%) for categorical variables and mean ± SD or median [IQR] for continuous variables.

Table 2. Parameter estimates and p-values for regression of plasma biomarkers on TMAO, choline, carnitine and betaine.

	TMAO		Choline		Carnitine		Betaine	
	β (SE)	p	β (SE)	p	β (SE)	p	β (SE)	p
CRP	0.01 (0.04)	0.86	**0.23 (0.10)**	**0.03**	-0.06 (0.12)	0.64	**0.19 (0.08)**	**0.02**
LBP	512.4 (283.9)	0.07	**1968.3 (795.8)**	**0.01***	1085.6 (921.5)	0.24	1131.0 (594.8)	0.06
HOMA-IR	**0.08 (0.03)**	**0.004***	**0.18 (0.08)**	**0.03**	**0.27 (0.09)**	**0.004***	**-0.32 (0.06)**	**<0.001***
HDL cholesterol	-1.49 (0.82)	0.07	**-14.86 (2.29)**	**<0.001***	**-9.64 (2.53)**	**<0.001***	0.23 (1.73)	0.86
LDL cholesterol	-1.47 (1.79)	0.41	**9.86 (5.01)**	**0.05**	-6.81 (6.03)	0.26	-0.21 (3.74)	0.96
Triglycerides	0.01 (0.02)	0.43	**0.14 (0.05)**	**0.002***	**0.18 (0.05)**	**<0.001***	**-0.16 (0.03)**	**<0. 001***
SBP	0.06 (0.62)	0.92	**4.50 (1.73)**	**0.01***	-2.33 (2.05)	0.24	**-2.68 (1.29)**	**0.04**
DBP	-0.35 (0.37)	0.34	1.33 (1.04)	0.20	-1.04 (1.22)	0.39	**-2.27 (0.78)**	**0.003***

Associations between TMAO, choline, and betaine with biomarkers were assessed in n=1716 participants, while carnitine and biomarker associations were assessed in n=1434 participants. Regression models were adjusted for age (continuous), sex, race-ethnicity (five groups), physical activity (METs), percent body fat (continuous), and aspirin use (no, previous, current).
*Adjusted p<0.05 after correcting for FDR by Benjamini-Hochberg.

Table 3. Associations between gut microbiome community structure and plasma TMAO, choline, carnitine, and betaine using PERMANOVA.

	Unweighted UniFrac		Weighted UniFrac	
	R^2	p	R^2	p
TMAO	**0.0010**	**0.007**	**0.0014**	**0.038**
Choline	**0.0010**	**0.006**	0.0012	0.068
Carnitine	**0.0018**	**0.002**	**0.0029**	**0.007**
Betaine	0.0008	0.073	0.0012	0.071

R^2 and p-values estimated after controlling for laboratory batch.

Table 4. Fecal bacterial genera associated with plasma TMAO from LASSO regression.

Phylum__Class__Order__Family__Genus	Prevalence	Coefficient
Positive association		
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae_Incertae_Sedis__Clostridia	71.8	0.0140
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__Blautia	100.0	0.0073
Firmicutes__Clostridia__Clostridiales__Peptostreptococcaceae_Incertae_Sedis__uncultured_bacterium	89.7	0.0004
Proteobacteria__Deltaproteobacteria__Desulfovibrionales__Desulfovibrionaceae__Desulfovibrio	24.7	0.0064
Negative association		
Bacteroidetes__Bacteroidia__Bacteroidales__Porphyromonadaceae__Barnesiella	100.0	-0.0198
Firmicutes__Clostridia__Clostridiales__Peptostreptococcaceae__uncultured	90.9	-0.0323
Firmicutes__Clostridia__Clostridiales__Ruminococcaceae__Faecalibacterium	100.0	-0.0240
Firmicutes__Clostridia__Clostridiales__uncultured__uncultured_bacterium	27.3	-0.0044
Firmicutes__Clostridia__Halanaerobiales__Halanaerobiaceae__Halocella	22.2	-0.0109
Firmicutes__Erysipelotrichi__Erysipelotrichales__Erysipelotrichaceae_Incertae_Sedis__Clostridium	58.1	-0.0173

Coefficients estimated after adjustment for laboratory batch. Prevalence indicates percent of participants with each bacterium.

Supplemental Table 1. Median (IQR) values of choline, carnitine, and betaine across quartiles of TMAO for all participants and across race-ethnic groups.

	TMAO Quartile			
	Q1 <2.10 µmol/L	Q2 2.10-3.05 µmol/L	Q3 3.06-4.60 µmol/L	Q4 >4.60 µmol/L
Choline (µmol/L)				
All participants	11.5 [9.9-13.3]	12.3 [10.6-14.2]	12.8 [10.8-14.9]	13.1 [11.0-15.8]
African American	11.9 [10.5-13.1]	12.3 [10.3-14.1]	13.2 [11.4-15.2]	14.0 [11.9-16.7]
Native Hawaiian	10.9 [9.7-12.5]	12.3 [11.0-14.1]	12.9 [11.1-15.0]	13.1 [10.5-16.3]
Japanese American	12.1 [10.0-14.2]	12.2 [10.6-14.3]	12.4 [10.6-15.9]	13.6 [11.6-16.2]
Latino	12.2 [10.7-13.5]	13.0 [11.2-15.9]	13.8 [11.7-15.5]	13.1 [11.4-15.9]
White	10.5 [9.2-11.6]	11.5 [10.3-13.3]	11.8 [10.1-13.7]	12.3 [10.2-14.3]
Carnitine (µmol/L)				
All participants	38.7 [32.7-44.3]	38.7 [32.9-44.0]	39.2 [34.2-45.5]	40.9 [35.6-46.8]
African American	33.7 [30.1-38.8]	38.0 [32.1-42.8]	36.7 [32.7-42.5]	39.6 [32.7-42.9]
Native Hawaiian	40.3 [37.2-45.7]	40.6 [36.6-46.8]	40.4 [35.5-46.1]	42.4 [37.0-47.5]
Japanese American	41.8 [36.9-47.6]	41.8 [36.8-47.4]	43.5 [37.6-49.5]	42.5 [38.6-50.2]
Latino	37.3 [32.8-44.2]	36.0 [31.3-43.2]	37.9 [32.7-45.0]	39.2 [32.5-47.0]
White	33.0 [28.9-39.7]	35.1 [31.3-41.0]	38.0 [32.5-42.9]	39.9 [35.2-45.3]
Betaine (µmol/L)				
All participants	39.9 [31.7-47.7]	40.3 [33.5-49.2]	41.2 [33.6-51.1]	40.4 [31.4-48.8]
African American	38.2 [30.9-45.8]	38.2 [31.2-47.0]	43.2 [32.3-51.5]	41.1 [31.8-47.8]
Native Hawaiian	41.3 [30.9-50.3]	44.0 [36.4-49.1]	43.2 [35.4-51.4]	42.3 [33.8-50.5]
Japanese American	41.9 [34.3-48.4]	42.2 [34.2-54.7]	42.3 [32.8-52.2]	44.1 [33.8-53.7]
Latino	39.8 [32.9-47.3]	39.3 [34.0-45.1]	39.9 [35.4-49.0]	39.0 [31.8-46.5]
White	36.1 [29.7-45.5]	38.9 [31.9-47.9]	39.3 [32.2-49.8]	35.9 [30.0-44.7]

Supplemental Table 2. Parameter estimates and p-values for regression of plasma biomarkers on TMAO, choline, and betaine after further adjustment for energy intake and dietary intake of seafood and red and processed meats.

	TMAO		Choline		Carnitine		Betaine	
	β (SE)	p	β (SE)	p	β (SE)	p	β (SE)	p
CRP	-0.01 (0.04)	0.95	**0.21 (0.10)**	**0.04**	-0.04 (0.12)	0.74	**0.21 (0.08)**	**0.01**
LBP	435.4 (284.2)	0.13	**1749.4 (794.7)**	**0.03**	941.5 (920.9)	0.31	**1241.1 (592.9)**	**0.04**
HOMA-IR	**0.07 (0.03)**	**0.01**	0.14 (0.08)	0.07	**0.24 (0.09)**	**0.01**	**-0.30 (0.06)**	**<0.001**
HDL cholesterol	-1.46 (0.83)	0.08	**-14.71 (2.29)**	**<0.001**	**-9.57 (2.53)**	**<0.001**	0.04 (1.73)	0.98
LDL cholesterol	-1.44 (1.80)	0.42	**10.08 (5.03)**	**0.05**	-6.76 (6.05)	0.26	-0.27 (3.75)	0.94
Triglycerides	0.01 (0.02)	0.45	**0.13 (0.05)**	**0.004**	**0.18 (0.05)**	**<0.001**	**-0.16 (0.03)**	**<0.001**
SBP	-0.04 (0.62)	0.94	**4.29 (1.73)**	**0.01**	-2.61 (2.05)	0.20	-2.53 (1.29)	0.05
DBP	-0.41 (0.37)	0.28	1.30 (1.04)	0.21	-1.10 (1.22)	0.37	**-2.19 (0.78)**	**0.005**

Associations between TMAO, choline, and betaine with biomarkers were assessed in n=1716 participants, while carnitine and biomarker associations were assessed in n=1434 participants.

Supplemental Table 3. Fecal bacterial genera associated with plasma choline from LASSO regression.

Phylum__Class__Order__Family__Genus	Prevalence	Coefficient
Positive association		
Bacteroidetes__Bacteroidia__Bacteroidales__Other__Other	90.8	0.0063
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae_Incertae_Sedis__Clostridia	71.8	0.0007
Firmicutes__Clostridia__Clostridiales__Peptostreptococcaceae_Incertae_Sedis__uncultured_bacterium	89.7	0.0084
Proteobacteria__Betaproteobacteria__Burkholderiales__Alcaligenaceae__Sutterella	83.9	0.0054
Proteobacteria__Gammaproteobacteria__Enterobacteriales__Enterobacteriaceae__Raoultella	74.5	0.0081
Negative association		
Firmicutes__Clostridia__Clostridiales__Christensenellaceae__uncultured	99.9	-0.0017
Firmicutes__Clostridia__Clostridiales__Lachnospiraceae__Shuttleworthia	43.6	-0.0002
Firmicutes__Clostridia__Clostridiales__Ruminococcaceae__Oscillospira	89.3	-0.0050
Firmicutes__Clostridia__Clostridiales__Ruminococcaceae__uncultured	100.0	-0.0004

Coefficients estimated after adjustment for laboratory batch. Prevalence indicates percent of participants with each bacterium.

Supplemental Table 4. Associations between TMAO and choline with *M. smithii*. Linear regression models were adjusted for age, sex, race-ethnicity, and percent body fat.

	TMAO		Choline	
	β (SE)	p	β (SE)	p
M. smithii (continuous)	-0.003 (0.003)	0.268	-0.001 (0.001)	0.289
M. smithii (presence)	**0.071 (0.034)**	**0.043**	0.009 (0.012)	0.471

CHAPTER 4
Conclusion

In this dissertation, we used the MEC to conduct population-based studies of the fecal microbiome. We first assessed the temporal variability of the fecal microbiome, showing most measures derived from 16S rRNA sequencing to be reliable over a two-year period, although taxa with very low abundance tended to be more variable. These results show that a single sample sufficiently captures a majority of fecal microbiome variation, although multiple samples or larger sample sizes may be needed if rare taxa are of particular interest. We next focused on the gut microbial metabolite TMAO, which we found to be associated with genera known to metabolize choline into TMA based on previous work using *in vitro* and bioinformatics approaches. TMAO and its precursors (choline, carnitine, and betaine) were also associated with several inflammatory and cardiometabolic markers. Of note, choline was associated with adverse health outcomes, while TMAO, choline, carnitine, and betaine were all associated with insulin resistance.

The results of our TMAO analysis in particular pave the way for at least two avenues of further research. First, while we used 16S rRNA sequencing to analyze genus-level data, other methods such as whole-genome shotgun metagenomic sequencing provide phylogenetic resolution at lower taxonomic levels. This approach would allow one to determine associations between TMAO and microbial species, which could shed light on which specific members of our identified genera are driving the associations. This is important as species even within the same genus can have a wide range of functional capacities. Second, the associations between insulin resistance across our four compounds suggest a possible role in diabetes risk. A nested case-control study of diabetes cases and matched controls within the MEC or another population-

based study using pre-diagnostic blood samples to measure TMAO and its precursors would be one approach to follow-up on this finding.

The decreasing cost of sequencing, along with the greater understanding of how microbes in the gut and other body sites impact states of health and disease, have helped lead the way for the incorporation of microbiome data into larger, population-based studies. Prospective studies will be particularly important for building on much of the cross-sectional research that has been conducted to date. The results of our two aims highlight the utility of using epidemiological approaches to better understand the human microbiome, along with areas that need further study. As with other biospecimens, there are many technical sources of variation that need to be considered when dealing with microbiome samples. While we have addressed temporal variation, other issues including shipping and storage of samples, data processing, and statistical analysis are all areas that can impact findings in a study. Additional efforts by groups such as the Microbiome Quality Control (MBQC) project will be crucial in determining the reproducibility and accuracy of these various methods, and how they impact data quality and comparability across different studies.

There has also been growing interest in understanding the functional capacity and metabolic activity of the microbiome. This requires methods beyond 16S rRNA sequencing, and include metagenomics and metatranscriptomics, as well as quantification of microbially-derived metabolites, as we have done with TMAO. In the gut, microbes are involved in the production of a vast range of compounds with important physiological roles. Additional studies are needed to identify specific bacteria involved in these processes and the impact of these metabolites on disease risk.

In conclusion, we used population-based approaches to characterize the stability of the fecal microbiome and to identify associations between TMAO, the fecal microbiome, and inflammatory and cardiometabolic risk biomarkers. As shown by this dissertation, population-based studies of the human microbiome complement and are important in building upon

laboratory and computational studies. Epidemiology will play a pivotal role in our understanding of this exciting and rapidly growing field.

References

1. Huttunen R, Syrjanen J. Obesity and the risk and outcome of infection. Int J Obes (Lond) 2013;37(3):333-40. doi: 10.1038/ijo.2012.62.
2. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 2010;464(7285):59-65. doi: 10.1038/nature08821.
3. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, et al. Human genetics shape the gut microbiome. Cell 2014;159(4):789-99. doi: 10.1016/j.cell.2014.09.053.
4. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D, et al. Population-level analysis of gut microbiome variation. Science 2016;352(6285):560-4. doi: 10.1126/science.aad3503.
5. Tedjo DI, Jonkers DM, Savelkoul PH, Masclee AA, van Best N, Pierik MJ, Penders J. The effect of sampling and storage on the fecal microbiota composition in healthy and diseased subjects. PLoS One 2015;10(5):e0126685. doi: 10.1371/journal.pone.0126685.
6. Hang J, Desai V, Zavaljevski N, Yang Y, Lin X, Satya RV, Martinez LJ, Blaylock JM, Jarman RG, Thomas SJ, et al. 16S rRNA gene pyrosequencing of reference and clinical samples and investigation of the temperature stability of microbiome profiles. Microbiome 2014;2:31. doi: 10.1186/2049-2618-2-31.
7. Wu GD, Lewis JD, Hoffmann C, Chen YY, Knight R, Bittinger K, Hwang J, Chen J, Berkowsky R, Nessel L, et al. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. BMC Microbiol 2010;10:206. doi: 10.1186/1471-2180-10-206.
8. Carroll IM, Ringel-Kulka T, Siddle JP, Klaenhammer TR, Ringel Y. Characterization of the fecal microbiota using high-throughput sequencing reveals a stable microbial

community during storage. PLoS One 2012;7(10):e46953. doi: 10.1371/journal.pone.0046953.

9. Lauber CL, Zhou N, Gordon JI, Knight R, Fierer N. Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. FEMS Microbiol Lett 2010;307(1):80-6. doi: 10.1111/j.1574-6968.2010.01965.x.

10. Voigt AY, Costea PI, Kultima JR, Li SS, Zeller G, Sunagawa S, Bork P. Temporal and technical variability of human gut metagenomes. Genome Biol 2015;16:73. doi: 10.1186/s13059-015-0639-8.

11. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S, Abnet CC, et al. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol 2017;35(11):1077-86. doi: 10.1038/nbt.3981.

12. Satinsky BM, Gifford SM, Crump BC, Moran MA. Use of internal standards for quantitative metatranscriptome and metagenome analysis. Methods Enzymol 2013;531:237-50. doi: 10.1016/B978-0-12-407863-5.00012-5.

13. Sinha R, Abnet CC, White O, Knight R, Huttenhower C. The microbiome quality control project: baseline study design and future directions. Genome Biol 2015;16(1):276. doi: 10.1186/s13059-015-0841-8.

14. Sinha R, Abnet CC, White O, Knight R, Huttenhower C. The microbiome quality control project: baseline study design and future directions. Genome Biol 2015;16:276. doi: 10.1186/s13059-015-0841-8.

15. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. Science 2009;326(5960):1694-7. doi: 10.1126/science.1177486.

16. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N, et al. Moving pictures of the human microbiome. Genome Biol 2011;12(5):R50. doi: 10.1186/gb-2011-12-5-r50.

17. Claesson MJ, Cusack S, O'Sullivan O, Greene-Diniz R, de Weerd H, Flannery E, Marchesi JR, Falush D, Dinan T, Fitzgerald G, et al. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. Proc Natl Acad Sci U S A 2011;108 Suppl 1:4586-91. doi: 10.1073/pnas.1000097107.

18. Flores GE, Caporaso JG, Henley JB, Rideout JR, Domogala D, Chase J, Leff JW, Vazquez-Baeza Y, Gonzalez A, Knight R, et al. Temporal variability is a personalized feature of the human microbiome. Genome Biol 2014;15(12):531. doi: 10.1186/s13059-014-0531-y.

19. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, et al. The long-term stability of the human gut microbiota. Science 2013;341(6141):1237439. doi: 10.1126/science.1237439.

20. Mehta RS, Abu-Ali GS, Drew DA, Lloyd-Price J, Subramanian A, Lochhead P, Joshi AD, Ivey KL, Khalili H, Brown GT, et al. Stability of the human faecal microbiome in a cohort of adult men. Nat Microbiol 2018;3(3):347-55. doi: 10.1038/s41564-017-0096-0.

21. den Besten G, van Eunen K, Groen AK, Venema K, Reijngoud DJ, Bakker BM. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. J Lipid Res 2013;54(9):2325-40. doi: 10.1194/jlr.R036012.

22. Singh RK, Chang HW, Yan D, Lee KM, Ucmak D, Wong K, Abrouk M, Farahnik B, Nakamura M, Zhu TH, et al. Influence of diet on the gut microbiome and implications for human health. J Transl Med 2017;15(1):73. doi: 10.1186/s12967-017-1175-y.

23. LeBlanc JG, Milani C, de Giori GS, Sesma F, van Sinderen D, Ventura M. Bacteria as vitamin suppliers to their host: a gut microbiota perspective. Curr Opin Biotechnol 2013;24(2):160-8. doi: 10.1016/j.copbio.2012.08.005.

24. Mardinoglu A, Shoaie S, Bergentall M, Ghaffari P, Zhang C, Larsson E, Backhed F, Nielsen J. The gut microbiota modulates host amino acid and glutathione metabolism in mice. Mol Syst Biol 2015;11(10):834. doi: 10.15252/msb.20156487.

25. Tang WH, Wang Z, Levison BS, Koeth RA, Britt EB, Fu X, Wu Y, Hazen SL. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. N Engl J Med 2013;368(17):1575-84. doi: 10.1056/NEJMoa1109400.

26. Lever M, George PM, Slow S, Bellamy D, Young JM, Ho M, McEntyre CJ, Elmslie JL, Atkinson W, Molyneux SL, et al. Betaine and Trimethylamine-N-Oxide as Predictors of Cardiovascular Outcomes Show Different Patterns in Diabetes Mellitus: An Observational Study. PLoS One 2014;9(12):e114969. doi: 10.1371/journal.pone.0114969.

27. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung YM, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 2011;472(7341):57-63. doi: 10.1038/nature09922.

28. Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, Britt EB, Fu X, Wu Y, Li L, et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat Med 2013;19(5):576-85. doi: 10.1038/nm.3145.

29. Bae S, Ulrich CM, Neuhouser ML, Malysheva O, Bailey LB, Xiao L, Brown EC, Cushing-Haugen KL, Zheng Y, Cheng TY, et al. Plasma choline metabolites and colorectal cancer risk in the Women's Health Initiative Observational Study. Cancer Res 2014;74(24):7442-52. doi: 10.1158/0008-5472.can-14-1835.

30. Mondul AM, Moore SC, Weinstein SJ, Karoly ED, Sampson JN, Albanes D. Metabolomic analysis of prostate cancer risk in a prospective cohort: The alpha-tocopherol, beta-carotene cancer prevention study. Int J Cancer 2015. doi: 10.1002/ijc.29576.

31. Shan Z, Sun T, Huang H, Chen S, Chen L, Luo C, Yang W, Yang X, Yao P, Cheng J, et al. Association between microbiota-dependent metabolite trimethylamine-N-oxide and type 2 diabetes. Am J Clin Nutr 2017;106(3):888-94. doi: 10.3945/ajcn.117.157107.

32. Tang WH, Wang Z, Fan Y, Levison B, Hazen JE, Donahue LM, Wu Y, Hazen SL. Prognostic value of elevated levels of intestinal microbe-generated metabolite trimethylamine-N-oxide in patients with heart failure: refining the gut hypothesis. J Am Coll Cardiol 2014;64(18):1908-14. doi: 10.1016/j.jacc.2014.02.617.

33. Troseid M, Ueland T, Hov JR, Svardal A, Gregersen I, Dahl CP, Aakhus S, Gude E, Bjorndal B, Halvorsen B, et al. Microbiota-dependent metabolite trimethylamine-N-oxide is associated with disease severity and survival of patients with chronic heart failure. J Intern Med 2014. doi: 10.1111/joim.12328.

34. Bain MA, Faull R, Fornasini G, Milne RW, Evans AM. Accumulation of trimethylamine and trimethylamine-N-oxide in end-stage renal disease patients undergoing haemodialysis. Nephrol Dial Transplant 2006;21(5):1300-4. doi: 10.1093/ndt/gfk056.

35. Tang WH, Wang Z, Kennedy DJ, Wu Y, Buffa JA, Agatisa-Boyle B, Li XS, Levison BS, Hazen SL. Gut Microbiota-Dependent Trimethylamine N-Oxide (TMAO) Pathway Contributes to Both Development of Renal Insufficiency and Mortality Risk in Chronic Kidney Disease. Circ Res 2015;116(3):448-55. doi: 10.1161/circresaha.116.305360.

36. Libby P. Inflammation in atherosclerosis. Nature 2002;420(6917):868-74. doi: 10.1038/nature01323.

37. Miller CA, Corbin KD, da Costa KA, Zhang S, Zhao X, Galanko JA, Blevins T, Bennett BJ, O'Connor A, Zeisel SH. Effect of egg ingestion on trimethylamine-N-oxide production in humans: a randomized, controlled, dose-response study. Am J Clin Nutr 2014;100(3):778-86. doi: 10.3945/ajcn.114.087692.

38. Kaysen GA, Johansen KL, Chertow GM, Dalrymple LS, Kornak J, Grimes B, Dwyer T, Chassy AW, Fiehn O. Associations of Trimethylamine N-Oxide With Nutritional and

Inflammatory Biomarkers and Cardiovascular Outcomes in Patients New to Dialysis. J Ren Nutr 2015. doi: 10.1053/j.jrn.2015.02.006.

39. Gao X, Liu X, Xu J, Xue C, Xue Y, Wang Y. Dietary trimethylamine N-oxide exacerbates impaired glucose tolerance in mice fed a high fat diet. J Biosci Bioeng 2014;118(4):476-81. doi: 10.1016/j.jbiosc.2014.03.001.

40. Rohrmann S, Linseisen J, Allenspach M, von Eckardstein A, Muller D. Plasma Concentrations of Trimethylamine-N-oxide Are Directly Associated with Dairy Food Consumption and Low-Grade Inflammation in a German Adult Population. J Nutr 2016;146(2):283-9. doi: 10.3945/jn.115.220103.

41. Seldin MM, Meng Y, Qi H, Zhu W, Wang Z, Hazen SL, Lusis AJ, Shih DM. Trimethylamine N-Oxide Promotes Vascular Inflammation Through Signaling of Mitogen-Activated Protein Kinase and Nuclear Factor-kappaB. J Am Heart Assoc 2016;5(2). doi: 10.1161/jaha.115.002767.

42. Schirra HJ, Anderson CG, Wilson WJ, Kerr L, Craik DJ, Waters MJ, Lichanska AM. Altered metabolism of growth hormone receptor mutant mice: a combined NMR metabonomics and microarray study. PLoS One 2008;3(7):e2764. doi: 10.1371/journal.pone.0002764.

43. Miao J, Ling AV, Manthena PV, Gearing ME, Graham MJ, Crooke RM, Croce KJ, Esquejo RM, Clish CB, Vicent D, et al. Flavin-containing monooxygenase 3 as a potential player in diabetes-associated atherosclerosis. Nat Commun 2015;6:6498. doi: 10.1038/ncomms7498.

44. Konstantinova SV, Tell GS, Vollset SE, Nygard O, Bleie O, Ueland PM. Divergent associations of plasma choline and betaine with components of metabolic syndrome in middle age and elderly men and women. J Nutr 2008;138(5):914-20. doi: 10.1093/jn/138.5.914.

45. Roe AJ, Zhang S, Bhadelia RA, Johnson EJ, Lichtenstein AH, Rogers GT, Rosenberg IH, Smith CE, Zeisel SH, Scott TM. Choline and its metabolites are differently associated with cardiometabolic risk factors, history of cardiovascular disease, and MRI-documented cerebrovascular disease in older adults. Am J Clin Nutr 2017;105(6):1283-90. doi: 10.3945/ajcn.116.137158.

46. Lever M, George PM, Atkinson W, Molyneux SL, Elmslie JL, Slow S, Richards AM, Chambers ST. Plasma lipids and betaine are related in an acute coronary syndrome cohort. PLoS One 2011;6(7):e21666. doi: 10.1371/journal.pone.0021666.

47. Tang WH, Hazen SL. The contributory role of gut microbiota in cardiovascular disease. J Clin Invest 2014;124(10):4204-11. doi: 10.1172/jci72331.

48. Lang DH, Yeung CK, Peter RM, Ibarra C, Gasser R, Itagaki K, Philpot RM, Rettie AE. Isoform specificity of trimethylamine N-oxygenation by human flavin-containing monooxygenase (FMO) and P450 enzymes: selective catalysis by FMO3. Biochem Pharmacol 1998;56(8):1005-12.

49. Chao CK, Zeisel SH. Formation of trimethylamine from dietary choline by Streptococcus sanguis I, which colonizes the mouth. J Nutr Biochem 1990;1(2):89-97.

50. Dyer FE, Wood A. Action of Enterobacteriaceae on choline and related compounds. J Fish Res Board Can 1947;7(1):17-21.

51. Ditullio D, Anderson D, Chen CS, Sih CJ. L-carnitine via enzyme-catalyzed oxidative kinetic resolution. Bioorg Med Chem 1994;2(6):415-20.

52. Unemoto T, Hayashi M, Miyaki K, Hayashi M. Formation of trimethylamine from DL-carnitine by Serratia marcescens. Biochim Biophys Acta 1966;121(1):220-2.

53. Romano KA, Vivas EI, Amador-Noguez D, Rey FE. Intestinal microbiota composition modulates choline bioavailability from diet and accumulation of the proatherogenic metabolite trimethylamine-N-oxide. MBio 2015;6(2). doi: 10.1128/mBio.02481-14.

54. Craciun S, Balskus EP. Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. Proc Natl Acad Sci U S A 2012;109(52):21307-12. doi: 10.1073/pnas.1215689109.

55. Zhu Y, Jameson E, Crosatti M, Schafer H, Rajakumar K, Bugg TD, Chen Y. Carnitine metabolism to trimethylamine by an unusual Rieske-type oxygenase from human microbiota. Proc Natl Acad Sci U S A 2014;111(11):4268-73. doi: 10.1073/pnas.1316569111.

56. Rath S, Heidrich B, Pieper DH, Vital M. Uncovering the trimethylamine-producing bacteria of the human gut microbiota. Microbiome 2017;5(1):54. doi: 10.1186/s40168-017-0271-9.

57. Sears CL, Garrett WS. Microbes, microbiota, and colon cancer. Cell Host Microbe 2014;15(3):317-28. doi: 10.1016/j.chom.2014.02.007.

58. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 2012;13(9):R79. doi: 10.1186/gb-2012-13-9-r79.

59. Sinha R, Chen J, Amir A, Vogtmann E, Shi J, Inman KS, Flores R, Sampson J, Knight R, Chia N. Collecting Fecal Samples for Microbiome Analyses in Epidemiology Studies. Cancer Epidemiol Biomarkers Prev 2016;25(2):407-16. doi: 10.1158/1055-9965.epi-15-0951.

60. Flores R, Shi J, Yu G, Ma B, Ravel J, Goedert JJ, Sinha R. Collection media and delayed freezing effects on microbial composition of human stool. Microbiome 2015;3:33. doi: 10.1186/s40168-015-0092-7.

61. Li F, Hullar MA, Lampe JW. Optimization of terminal restriction fragment polymorphism (TRFLP) analysis of human gut microbiota. J Microbiol Methods 2007;68(2):303-11. doi: 10.1016/j.mimet.2006.09.006.

62. Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS One 2012;7(2):e30087. doi: 10.1371/journal.pone.0030087.

63. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One 2011;6(12):e27310. doi: 10.1371/journal.pone.0027310.

64. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods 2013;10(1):57-9. doi: 10.1038/nmeth.2276.

65. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science 2011;334(6052):105-8. doi: 10.1126/science.1208344.

66. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature 2014;505(7484):559-63. doi: 10.1038/nature12820.

67. Li F, Hullar MA, Schwarz Y, Lampe JW. Human gut bacterial communities are altered by addition of cruciferous vegetables to a controlled fruit- and vegetable-free diet. J Nutr 2009;139(9):1685-91. doi: 10.3945/jn.109.108191.

68. O'Keefe SJ, Li JV, Lahti L, Ou J, Carbonero F, Mohammed K, Posma JM, Kinross J, Wahl E, Ruder E, et al. Fat, fibre and cancer risk in African Americans and rural Africans. Nat Commun 2015;6:6342. doi: 10.1038/ncomms7342.

69. Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci U S A 2011;108 Suppl 1:4554-61. doi: 10.1073/pnas.1000087107.

70. Chang JY, Antonopoulos DA, Kalra A, Tonelli A, Khalife WT, Schmidt TM, Young VB. Decreased diversity of the fecal Microbiome in recurrent Clostridium difficile-associated diarrhea. J Infect Dis 2008;197(3):435-8. doi: 10.1086/525047.

71. Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, Stram DO, Monroe KR, Earle ME, Nagamine FS. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol 2000;151(4):346-57.

72. Maskarinec G, Lim U, Jacobs S, Monroe KR, Ernst T, Buchthal SD, Shepherd JA, Wilkens LR, Le Marchand L, Boushey CJ. Diet Quality in Midadulthood Predicts Visceral Adiposity and Liver Fatness in Older Ages: The Multiethnic Cohort Study. Obesity (Silver Spring) 2017;25(8):1442-50. doi: 10.1002/oby.21868.

73. Fu BC, Randolph TW, Lim U, Monroe KR, Cheng I, Wilkens LR, Le Marchand L, Hullar MA, Lampe JW. Characterization of the gut microbiome in epidemiologic studies: the multiethnic cohort experience. Ann Epidemiol 2016;26(5):373-9. doi: 10.1016/j.annepidem.2016.02.009.

74. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7(5):335-6. doi: 10.1038/nmeth.f.303.

75. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010;26(19):2460-1. doi: 10.1093/bioinformatics/btq461.

76. Navas-Molina JA, Peralta-Sanchez JM, Gonzalez A, McMurdie PJ, Vazquez-Baeza Y, Xu Z, Ursell LK, Lauber C, Zhou H, Song SJ, et al. Advancing our understanding of the human microbiome using QIIME. Methods Enzymol 2013;531:371-444. doi: 10.1016/b978-0-12-407863-5.00019-8.

77. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215(3):403-10. doi: 10.1016/s0022-2836(05)80360-2.

78. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 2009;75(23):7537-41. doi: 10.1128/aem.01541-09.

79. 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(16):5261-7. doi: 10.1128/aem.00062-07.

80. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 2007;35(21):7188-96. doi: 10.1093/nar/gkm864.

81. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 2010;26(2):266-7. doi: 10.1093/bioinformatics/btp636.

82. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 2010;5(3):e9490. doi: 10.1371/journal.pone.0009490.

83. Faith DP. Conservation evaluation and phylogenetic diversity. Biol Conserv 1992;61(1):1-10.

84. Shannon CE, Weaver W. The mathematical theory of communication: University of Illinois press, 1998.

85. Chao A. Nonparametric estimation of the number of classes in a population. Scand J Stat 1984:265-70.

86. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 2005;71(12):8228-35. doi: 10.1128/aem.71.12.8228-8235.2005.

87. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol 2007;73(5):1576-85. doi: 10.1128/aem.01996-06.

88. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci 2003;14(6):927-30.

89. Wu JR, Macklaim JM, Genge BL, Gloor GB. Finding the centre: corrections for asymmetry in high-throughput sequencing datasets. ArXiv e-prints 2017;1704.

90. Aitchison J. The statistical analysis of compositional data. J R Stat Soc Series B Stat Methodol 1982:139-77.

91. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. 2015 2015;67(1):48. doi: 10.18637/jss.v067.i01.

92. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6(4):284.

93. Sze MA, Schloss PD. Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome. MBio 2016;7(4). doi: 10.1128/mBio.01018-16.

94. Ley RE. Obesity and the human microbiome. Curr Opin Gastroenterol 2010;26(1):5-11. doi: 10.1097/MOG.0b013e328333d751.

95. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, et al. A core gut microbiome in obese and lean twins. Nature 2009;457(7228):480-4. doi: 10.1038/nature07540.

96. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, Barnes R, Watson P, Allen-Vercoe E, Moore RA, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res 2012;22(2):299-306. doi: 10.1101/gr.126516.111.

97. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, Clancy TE, Chung DC, Lochhead P, Hold GL, et al. Fusobacterium nucleatum potentiates intestinal

tumorigenesis and modulates the tumor-immune microenvironment. Cell Host Microbe 2013;14(2):207-15. doi: 10.1016/j.chom.2013.07.007.

98. Jakobsson HE, Jernberg C, Andersson AF, Sjolund-Karlsson M, Jansson JK, Engstrand L. Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. PLoS One 2010;5(3):e9836. doi: 10.1371/journal.pone.0009836.

99. Dethlefsen L, Huse S, Sogin ML, Relman DA. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS Biol 2008;6(11):e280. doi: 10.1371/journal.pbio.0060280.

100. O'Sullivan O, Coakley M, Lakshminarayanan B, Conde S, Claesson MJ, Cusack S, Fitzgerald AP, O'Toole PW, Stanton C, Ross RP. Alterations in intestinal microbiota of elderly Irish subjects post-antibiotic therapy. J Antimicrob Chemother 2013;68(1):214-21. doi: 10.1093/jac/dks348.

101. Cox LM, Blaser MJ. Antibiotics in early life and obesity. Nat Rev Endocrinol 2015;11(3):182-90. doi: 10.1038/nrendo.2014.210.

102. Bailey LC, Forrest CB, Zhang P, Richards TM, Livshits A, DeRusso PA. Association of antibiotics in infancy with early childhood obesity. JAMA Pediatr 2014;168(11):1063-9. doi: 10.1001/jamapediatrics.2014.1539.

103. Kohanski MA, Dwyer DJ, Collins JJ. How antibiotics kill bacteria: from targets to networks. Nat Rev Microbiol 2010;8(6):423-35. doi: 10.1038/nrmicro2333.

104. Zaura E, Brandt BW, Teixeira de Mattos MJ, Buijs MJ, Caspers MP, Rashid MU, Weintraub A, Nord CE, Savell A, Hu Y, et al. Same Exposure but Two Radically Different Responses to Antibiotics: Resilience of the Salivary Microbiome versus Long-Term Microbial Shifts in Feces. MBio 2015;6(6):e01693-15. doi: 10.1128/mBio.01693-15.

105. Dambrova M, Latkovskis G, Kuka J, Strele I, Konrade I, Grinberga S, Hartmane D, Pugovics O, Erglis A, Liepinsh E. Diabetes is Associated with Higher Trimethylamine N-

oxide Plasma Levels. Exp Clin Endocrinol Diabetes 2016;124(4):251-6. doi: 10.1055/s-0035-1569330.

106. Missailidis C, Hallqvist J, Qureshi AR, Barany P, Heimburger O, Lindholm B, Stenvinkel P, Bergman P. Serum Trimethylamine-N-Oxide Is Strongly Related to Renal Function and Predicts Outcome in Chronic Kidney Disease. PLoS One 2016;11(1):e0141738. doi: 10.1371/journal.pone.0141738.

107. Stubbs JR, House JA, Ocque AJ, Zhang S, Johnson C, Kimber C, Schmidt K, Gupta A, Wetmore JB, Nolin TD. Serum trimethylamine-N-oxide is elevated in CKD and correlates with coronary atherosclerosis burden. J Am Soc Nephrol 2016;27(1):305-13.

108. Cho CE, Taesuwan S, Malysheva OV, Bender E, Tulchinsky NF, Yan J, Sutter JL, Caudill MA. Trimethylamine-N-oxide (TMAO) response to animal source foods varies among healthy young men and is influenced by their gut microbiota composition: A randomized controlled trial. Mol Nutr Food Res 2017;61(1). doi: 10.1002/mnfr.201600324.

109. Fennema D, Phillips IR, Shephard EA. Trimethylamine and Trimethylamine N-Oxide, a Flavin-Containing Monooxygenase 3 (FMO3)-Mediated Host-Microbiome Metabolic Axis Implicated in Health and Disease. Drug Metab Dispos 2016;44(11):1839-50. doi: 10.1124/dmd.116.070615.

110. Rebouche CJ, Seim H. Carnitine metabolism and its regulation in microorganisms and mammals. Annu Rev Nutr 1998;18:39-61. doi: 10.1146/annurev.nutr.18.1.39.

111. Brugere JF, Borrel G, Gaci N, Tottey W, O'Toole PW, Malpuech-Brugere C. Archaebiotics: proposed therapeutic use of archaea to prevent trimethylaminuria and cardiovascular disease. Gut Microbes 2014;5(1):5-10. doi: 10.4161/gmic.26749.

112. Stram DO, Hankin JH, Wilkens LR, Pike MC, Monroe KR, Park S, Henderson BE, Nomura AM, Earle ME, Nagamine FS, et al. Calibration of the dietary questionnaire for a multiethnic cohort in Hawaii and Los Angeles. Am J Epidemiol 2000;151(4):358-70.

113. Lim U, Ernst T, Buchthal SD, Latch M, Albright CL, Wilkens LR, Kolonel LN, Murphy SP, Chang L, Novotny R, et al. Asian women have greater abdominal and visceral adiposity than Caucasian women with similar body mass index. Nutr Diabetes 2011;1:e6. doi: 10.1038/nutd.2011.2.

114. 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 1972;18(6):499-502.

115. 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 1985;28(7):412-9.

116. Mojiminiyi OA, Abdella NA. Effect of homeostasis model assessment computational method on the definition and associations of insulin resistance. Clin Chem Lab Med 2010;48(11):1629-34. doi: 10.1515/cclm.2010.303.

117. Wang Z, Levison BS, Hazen JE, Donahue L, Li XM, Hazen SL. Measurement of trimethylamine-N-oxide by stable isotope dilution liquid chromatography tandem mass spectrometry. Anal Biochem 2014;455:35-40. doi: 10.1016/j.ab.2014.03.016.

118. Lenky CC, McEntyre CJ, Lever M. Measurement of marine osmolytes in mammalian serum by liquid chromatography-tandem mass spectrometry. Anal Biochem 2012;420(1):7-12. doi: 10.1016/j.ab.2011.09.013.

119. Baker GC, Smith JJ, Cowan DA. Review and re-analysis of domain-specific 16S primers. J Microbiol Methods 2003;55(3):541-55.

120. DeLong EF. Archaea in coastal marine environments. Proc Natl Acad Sci U S A 1992;89(12):5685-9.

121. Zeileis A, Kleiber C, Jackman S. Regression models for count data in R. J Stat Softw 2008;27(8):1-25.

122. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H. Package 'vegan'. Community ecology package, version 2013;2(9).

123. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33(1):1-22.

124. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological) 1995:289-300.

125. Steiber A, Kerner J, Hoppel CL. Carnitine: a nutritional, biosynthetic, and functional perspective. Mol Aspects Med 2004;25(5-6):455-73. doi: 10.1016/j.mam.2004.06.006.

126. Zeisel SH, da Costa KA. Choline: an essential nutrient for public health. Nutr Rev 2009;67(11):615-23. doi: 10.1111/j.1753-4887.2009.00246.x.

127. Zeisel SH, Mar MH, Howe JC, Holden JM. Concentrations of choline-containing compounds and betaine in common foods. J Nutr 2003;133(5):1302-7. doi: 10.1093/jn/133.5.1302.

128. Li Y, Wang DD, Chiuve SE, Manson JE, Willett WC, Hu FB, Qi L. Dietary phosphatidylcholine intake and type 2 diabetes in men and women. Diabetes Care 2015;38(2):e13-4. doi: 10.2337/dc14-2093.

129. Meyer KA, Benton TZ, Bennett BJ, Jacobs DR, Jr., Lloyd-Jones DM, Gross MD, Carr JJ, Gordon-Larsen P, Zeisel SH. Microbiota-Dependent Metabolite Trimethylamine N-Oxide and Coronary Artery Calcium in the Coronary Artery Risk Development in Young Adults Study (CARDIA). J Am Heart Assoc 2016;5(10). doi: 10.1161/jaha.116.003970.

130. Wang C, Harris WS, Chung M, Lichtenstein AH, Balk EM, Kupelnick B, Jordan HS, Lau J. n-3 Fatty acids from fish or fish-oil supplements, but not alpha-linolenic acid, benefit cardiovascular disease outcomes in primary- and secondary-prevention studies: a systematic review. Am J Clin Nutr 2006;84(1):5-17. doi: 10.1093/ajcn/84.1.5.

131. Iso H, Sato S, Folsom AR, Shimamoto T, Terao A, Munger RG, Kitamura A, Konishi M, Iida M, Komachi Y. Serum fatty acids and fish intake in rural Japanese, urban Japanese, Japanese American and Caucasian American men. Int J Epidemiol 1989;18(2):374-81.

132. Dragsted LO. Biomarkers of meat intake and the application of nutrigenomics. Meat Sci 2010;84(2):301-7. doi: 10.1016/j.meatsci.2009.08.028.

133. Borrel G, McCann A, Deane J, Neto MC, Lynch DB, Brugere JF, O'Toole PW. Genomics and metagenomics of trimethylamine-utilizing Archaea in the human gut microbiome. ISME J 2017;11(9):2059-74. doi: 10.1038/ismej.2017.72.