Published in final edited form as:

Ann Epidemiol. 2016 May ; 26(5): 373–379. doi:10.1016/j.annepidem.2016.02.009.

Characterization of the Gut Microbiome in Epidemiologic Studies: The Multiethnic Cohort Experience

Benjamin C. Fu^{a,b}, Timothy W. Randolph^a, Unhee Lim^c, Kristine R. Monroe^d, Iona Cheng^{e,f}, Lynne R. Wilkens^c, Loïc Le Marchand^c, Meredith A.J. Hullar^{a,*}, and Johanna W. Lampe^{a,b,*}

^aPublic Health Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, PO Box 19024, Seattle, WA 98109

^bDepartment of Epidemiology, University of Washington, Seattle, WA 98195

^cEpidemiology Program, University of Hawaii Cancer Center, University of Hawaii at Manoa, Honolulu, HI 96813

^dUniversity of Southern California, Los Angeles, CA 90033

^eCancer Prevention Institute of California, Fremont, CA 94538

^fStanford Cancer Institute, Stanford, CA 94305

Abstract

Purpose—Development of next-generation sequencing and accompanying bioinformatics tools has revolutionized characterization of microbial communities. As interest grows in the role of the human microbiome in health and disease, so does the need for well-powered, robustly designed epidemiologic studies. Here, we discuss sources of bias that can arise in gut microbiome research.

Methods—Research comparing methods of specimen collection, preservation, processing, and analysis of gut microbiome samples is reviewed. While selected studies are primarily based on the gut, many of the same principles are applicable to samples derived from other anatomical sites. Methods for participant recruitment and sampling of the gut microbiome implemented in an ongoing population-based study, the Multiethnic Cohort (MEC), are also described.

Results—Variation in methodologies can influence the results of human microbiome studies. To help minimize bias, techniques such as sample homogenization, addition of internal standards, and quality filtering should be adopted in protocols. Within the MEC, participant response rates to stool sample collection were comparable to other studies, and in-home stool sample collection yields sufficient high-quality DNA for gut microbiome analysis.

Conclusion—Application of standardized and quality controlled methods in human microbiome studies is necessary to ensure data quality and comparability among studies.

Corresponding author: Johanna W. Lampe, PhD, RD, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, M4-B402, Seattle, WA 98109, Tel: 1-206-667-6580, ; Email: jlampe@fredhutch.org

*these authors contributed equally, shared senior authorship

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

microbiome; gut bacteria; stool; Multiethnic Cohort

INTRODUCTION

Spurred by the development and implementation of novel sequencing technologies and bioinformatic techniques, the last decade has seen great strides in human microbiome research. Furthermore, substantial reductions in sequencing costs have facilitated incorporation of microbiome research into large-scale epidemiologic studies. Major efforts to characterize microbial communities [1,2] have helped elucidate the rich and diverse microbial landscape in and on the human body, as well as the substantial variation across individuals.

In addition to identifying which microbes are present, the functional capacity of the microbiome can be characterized with metagenomic sequencing which provides a snapshot of the genetic composition of microbial genomes. Bacteria respond rapidly to changes in their environment not only in abundance but also in composition of the metabolically active fraction of the microbial community. Metatranscriptomics (i.e., microbial rRNA or mRNA sequencing) can be used to measure ribosomal and actively transcribed messenger RNA to gain insight into gene expression patterns. While still fairly new, the utility of these methods continues to grow as more genes become annotated. Metaproteomic (proteins from the microbial community) and metabolomic approaches are also increasingly used as a way to examine the products of microbial metabolism [3]. The bacterial component chosen to represent the microbiome (DNA, RNA, protein, or metabolite) needs to be considered early as it will influence the experimental design and timing of metadata gathered from the human population.

With the increasing wealth of microbiome data being generated from these techniques, there is a need to better understand clinically meaningful differences, confounding factors, and causality when studying the microbiome in relation to disease risk. Epidemiology is well equipped to tackle these issues by building on the knowledge and analytical tools that have been developed in assessing microbial exposures and multifactorial diseases. Here, we provide a review of potential sources of variation and bias that can be introduced and should be considered when studying the gut microbiome in large, prospective, population-based studies, as well as describe some of our experience collecting fecal samples as part of a Multiethnic Cohort (MEC) study in Hawaii and California. Although our focus is on the gut microbial community (GMC) and its function, aspects of sample collection, preservation, processing, and analysis are also relevant to other anatomical sites.

WHERE TO SAMPLE THE GUT MICROBIOME

Studies of the gut microbiome often use stool samples, for which collection is non-invasive and can be carried out privately by study participants. Biopsies are a second option, especially in studies of colorectal cancer risk or other situations when a colonoscopy or sigmoidoscopy is indicated. However, the gut mucosal and luminal microbiomes are not

necessarily comparable [4,5]; several studies have shown differences between stool and biopsy samples [6–9], and microbial populations have been reported to differ also in biopsies collected at various locations along the gastrointestinal (GI) tract [6,10,11]. Swabs have been used as well, although both rectal swabs and swabs of fresh stool may differ from the previous two methods in terms of microbial composition and DNA yield [12–14].

Two other less common collection methods for interrogating the fecal and mucosal microbiomes, respectively, are fecal occult blood tests (FOBT) and formalin fixed paraffin-embedded (FFPE) tissue from surgical resection or biopsy. Stool collected with FOBT cards was reported to be similar to that collected directly into a storage tube, either with or without an RNA-stabilizing agent, in terms of microbial community structure and taxa distribution [15]. Another recent study found FOBT to have optimal stability and reproducibility compared to seven fecal sampling methods [16]. FFPE is often collected in clinical settings and has been used for bacterial identification in a variety of diseases [17–19], but faces sample quality issues related to low DNA yields, fragmentation, and sequence artifacts [20,21].

Sampling method is an important consideration in understanding disease etiology, as disease states may have varying effects on microbial communities depending on anatomical site. For example, Bajaj et al. reported no differences in the stool microbiota of patients with and without hepatic encephalopathy, while significant differences in microbes associated with colonic mucosal biopsies were present between the two groups [8]. Ultimately, the choice of sampling approach should be driven by the hypothesis being tested. For example, study of direct effects of microbes on the gut mucosa likely warrants use of colon tissue for characterization of the gut microbiome, whereas studies of microbial metabolism of dietary constituents may be better served by analysis of luminal contents or stool.

SAMPLING THE STOOL MICROBIOME

Clinical analysis of stool is commonplace [22,23], and collection methods can be readily applied to research settings. Various groups, including the Human Microbiome Project (HMP) [24], have developed protocols that allow participants to collect a stool sample in the privacy of their own homes. An issue with using stool is that the considerable variation in the gut environment may lead to an uneven distribution of microbes in the stool sample. The spatial distribution of gut microbes, both longitudinally and radially, is influenced by factors such as increasing pH levels from the proximal to distal colon [25] and higher concentrations of oxygen near the mucosa relative to the lumen [26]. Homogenization of whole stool is one approach for obtaining a more uniform sample; this has been shown to reduce the variation in both the amount of DNA extracted [27] and the relative abundance of bacterial taxa [28,29].

STOOL COLLECTION AND TRANSPORT

Several factors related to specimen handling may also influence the quantity and quality of nucleic acids present in the stool samples and therefore impact the microbiome data generated.

Temperature

Studies have investigated the effect of temperature either during sample transport or storage on microbial community structure, finding variation due to temperature to be less than that due to inter-individual differences [13,30,31]. However, storing samples at room temperature for more than 24 hours without preservative can have significant effects on bacterial community composition and RNA fragmentation [32,33]. In terms of specific microbes, Tedjo et al. [13] found no taxa to be associated with storage method, while Rubin et al. [34] reported that only one of 2781 taxa significantly differed across three temperatures (−20°C, 4°C, room temperature). Gorzelak et al. [29] observed changes in Firmicutes levels within 3 days and Bacteroidetes by 14 days for storage at −20°C, while Fouhy et al. [35] showed significant differences in two genera (*Faecalibacterium*, *Leuconostoc*) when comparing fresh and flash-frozen samples.

Storage Solution

Submersion of samples in a nucleic acid storage solution also aids in preservation and greatly adds to the ease of sampling in the home environment. The product commonly used for human microbiome studies is RNAlater® (Ambion). Several studies have shown RNAlater® to be an effective storage reagent for preserving RNA or DNA [36–38]. In cluster analyses, comparing frozen samples to those in RNAlater® only, samples nearly always grouped by individual when either DNA- or RNA- based methods were used [15,33,38,39]. However, some studies have found RNAlater® to reduce yield and purity of bacterial DNA [15,29,40], and one found alterations in several bacterial phyla over 72 hours of storage compared to frozen samples [41].

LABORATORY CONSIDERATIONS

Storage Time

For large cohort studies, the capacity to store specimens without extensive processing is an important cost consideration. Studies have found inter-individual variation in microbial composition to be greater than variation due to storage time regardless of storage length. Clustering of repeated samples within individual has been seen over periods of up to 14 days [42], 6 months [43], and over 2 years [38]. Although few studies have assessed changes in individual taxa over storage time, Lauber et al. [42] have reported that the relative abundance of some microbes significantly changed over two weeks across several storage temperatures although overall composition did not. While storage time appears to have minimal effect on the overall community structure, the effect on specific taxa requires further investigation.

DNA and RNA Extraction

Nucleic acid extraction procedures developed for eukaryotes cannot be used to efficiently extract nucleic acids from bacterial cells as the bacterial cell structure varies. Several different extraction methods are used, including kits specifically designed to isolate DNA or RNA from environmental samples. Studies have found differences in taxa and DNA yields between extraction methods [27,31,40,44,45]. Protocols that incorporate additional lysis steps typically provide the best results [27,46,47]. Mechanical disruption in combination

with chemical or enzymatic lysis in particular has been shown to reflect more accurately community structure and improve nucleic acids yields for both DNA and RNA [27,30,37,48–51].

Another source of technical bias in RNA analysis is associated with removing the majority of rRNA from a sample in order to enrich the mRNA prior to sequence analysis [38,51,52]. Addition of internal standards can help to normalize these biases [51]; however, appropriate standards for microbiome studies is an on-going discussion [53,54].

SEQUENCING AND DATA PROCESSING

Targeted amplicon sequencing and whole genome shotgun sequencing (WGS) are both used to profile microbial community structure. Targeted sequencing is more common today and typically uses the 16S rRNA gene, which contains highly conserved regions interspersed with nine hypervariable regions (commonly denoted as V1–V9). As high-throughput platforms (reviewed in [55]) are currently limited by amplicon length, one or a subset of these regions is used for taxonomic classification. The choice of region can affect results. V6–V9 has been found to have a significantly higher error rate compared to the V1–V3 and V3–V5 regions [56,57], and Chakravorty et al. [58] reported V2 and V3 to be the best at distinguishing between 110 pathogenic bacterial species separately at the genus level. Although no region appears to work consistently well across all bacteria, combining multiple regions seems to provide the most accuracy.

Unlike targeted sequencing, the increasingly popular WGS [59] sequences all DNA in a sample, providing insight not only into taxonomy but biological function as well. Many methods are already available for gene prediction and functional annotation to obtain these functional profiles [59,60]. However, WGS does introduce challenges such as sequencing of host DNA and more complicated and computationally intensive analysis.

While the use of next generation sequencing technologies allows for a more comprehensive survey of microbial communities, there are several other issues inherent in sequencing-based methods, including polymerase chain reaction (PCR) amplification [61–63], library formation [64], primer selection [49,65], 16S rRNA gene copy number [66,67], sequencing platform [68,69], and sampling depth [61,70]. Additionally, the probability of accurately calling the correct base decreases from the 5' to 3' end of a sequence due to sequencing chemistry. Inclusion of positive and negative control samples during sequencing aid in identification of resulting sequencing errors and chimeras (products of multiple partial sequences) along with any sample contamination [71]. Quality filtering, which includes assuring appropriate phred quality scores, minimum number of consecutive high quality reads, and OTU abundance threshold, also needs to be applied to reduce spurious OTUs [72,73]. OTUs that are present in very low abundance are often filtered out as they generally represent technical artifacts [72]. Several open-source packages such as QIIME [74] and mothur [75] can implement these bioinformatics methods prior to sequence alignment and taxonomic assignment.

In order to identify microbes, sequences are compared to a reference database. The operational taxonomic unit (OTU) has been coined for sequences that cluster on similarity. For example, a 97% measure of similarity between sequences is often considered sufficient for species assignment [76]. A reference database (e.g., SILVA, Greengenes, RDP; [77–79]) is then used for phylogenetic classification. Three approaches are used to identify sequences in a sample and estimate diversity and community composition: the closed approach only includes samples that match a reference sequence in a database and removes unknowns; the de novo approach clusters samples without an external reference; and the open approach, which first uses the closed approach to match sequences in a reference database and then clusters the sequences that didn't match into similarity based clusters [80,81]. This last approach is robust and allows discovery of new bacteria and can be parallelized for rapid processing of large datasets.

STATISTICAL ANALYSIS

In preparation for statistical analysis, normalization of sequence counts is required due to technical differences in sequencing that result in a variable number of sequence counts per sample. Normalization methods acknowledge that true sequence counts are not observed and taxon abundances are naturally viewed as relative percentages of the entire GMC. Traditionally, total sum scaling (i.e., relative abundance) has been used to normalize samples. However, the inherent autocorrelation of the resulting relative quantities may misrepresent the correlation between the true abundances in the microbial community [82,83]. Alternatively, the set of sequences in each sample might be randomly sub-sampled or “rarefied” to obtain a fixed count common to all samples [80,84]. A variety of normalization methods have been proposed including the use of internal standards and variations on those designed for RNA-seq analysis [51,64], but this remains an area of active research [85,86]. The limitations of the normalized as compared to unconstrained multivariate data should be acknowledged in any statistical analysis [86–88].

Statistical analysis may be aimed at various levels. At the individual taxon level, univariate models aim to detect group differences in taxon abundance or associations with a continuous or categorical outcome such as subject phenotype. Tests such as Wilcoxon, ANOVA or Kruskal-Wallis [89] have been used, as well as univariate regression models. An additional consideration is the existence of many rare taxa that are not observed in a large number of subjects resulting in a large number of zeros. Therefore, several statistical models specific to single taxon-level analysis have been proposed to account for this over-dispersion [64,84,87,90].

A limitation of modeling one taxon at a time is the inability to address larger scientific goals related to how microbes interact as part of a community. Consequently, concepts from ecology are central to many methods used to analyze these data [91]. Community-level analysis requires a multivariate statistical method, the most common being principal coordinates analysis (PCoA) and non-metric multidimensional scaling (NMDS) [91,92]. This allows the analysis to incorporate a variety of ecologically-based or phylogenetically-informed definitions of between-sample distance [93,94]. The UniFrac distance [95] (or a related distance [96]) is popular as it is based on the premise that taxa which share a large

fraction of branches in a reference phylogenetic tree are more similar than those sharing a small fraction; the unweighted UniFrac distance quantifies the presence/absence of taxa between two samples and simply counts the number of branches shared by both. Our analyses of microbiome data in the MEC have employed both UniFrac and Jensen-Shannon, while other common dissimilarity measures include Bray-Curtis and Jaccard. Additional methods that incorporate phylogenetic structure into PCoA are also available [97–99].

PCoA and NMDS plots provide graphical insight on relationships between GMC vectors and a phenotype of interest, but do not quantify potential associations. For this, multivariate tests may be carried out using nonparametric MANOVA [100], or the kernel-based regression association test of MiRKAT [101]. Importantly, these methods allow for models that adjust for additional covariates such as gender and age.

Finally, while these tests for association are fundamental, they do not reveal which taxa (or sub-communities) are associated with an outcome. We have therefore worked toward developing multivariate regression methods for inferring potential associations of GMC members with a scalar outcome. This is based on well-known methods for penalized regression with high-dimensional predictors, such as ridge regression and least absolute shrinkage and selection operator (LASSO) [102], in which we additionally account for phylogenetic structure among taxa as well as non-Euclidean dissimilarity measures using a kernel-based framework [103].

RECRUITMENT ISSUES RELATED TO STOOL SAMPLE COLLECTION

As with any epidemiologic study, it is important to anticipate both participant response rate and representability prior to conducting a population-based microbiome study. In a previous study of healthy, post-menopausal women, Feigelson et al. [104] found that only 59 of 300 eligible women (20%) returned both a stool sample and study questionnaire. However, those who enrolled were not significantly different from those who refused or were unable to be contacted, and compliance among those who tentatively agreed and received the kit was 80%. One potential reason for refusal was that samples had to be collected and shipped on the same day and only on certain days of the week, which the authors suggest could be remedied by allowing participants to store specimens in the freezer for shipment when convenient. In contrast, a biospecimen collection feasibility study of 351 adult participants in the German National Cohort, which allowed for shipment of stool samples on a later date, reported an 86% response rate and 98% compliance [105]. Collection of stool for infectious intestinal diseases had response rates near 40% in two larger population-based studies [106,107].

INTRA-INDIVIDUAL VARIATION IN GUT MICROBIAL PROFILES OVER TIME

Another concern in prospective cohort studies is whether a sample collected at a single time point can adequately characterize an individual's microbiome or whether multiple measures are necessary. Microorganisms have relatively short generation times, on a scale of hours or days [108], which allows for the potential of rapid changes in community structure.

Although the adult gut microbiome is considerably more stable than the infant microbiome [70,109–112], many factors may influence its composition both in the short- and long-term, including diet [113–116], antibiotic use [117–119], and weight loss [110,120,121]. Variation is also seen in metabolite phenotype studies, which show that although phenotypes tend to be highly concordant over several years, bacterial capacity to metabolize certain dietary bioactives can change [122]. Thus, these data would suggest that the gut microbiome characterized using bacterial DNA is likely to be representative within an individual in comparison to other members of a cohort over days to several years.

THE MEC EXPERIENCE

As part of an ongoing human genome-wide association study (GWAS), we have collected a stool sample by mail from a subset of the Multiethnic Cohort (MEC) participants (age 60–90 years), recruited in Hawaii and California. These participants, who were previously genotyped using various SNP arrays as controls in prior GWAS, received an invitation letter in the mail, followed by a phone call from study recruiters. If interested, they answered screening questions to determine immediate or deferred eligibility: no ileostomy or colectomy; not on dialysis; no recent treatment with chemotherapy, radiation therapy, corticosteroid hormones, prescription weight-loss drugs, insulin or thyroid medications (<6 months); no recent endoscopy or irrigation/cleansing of the large intestine (<6 months); no recent antibiotics in oral or IV form (<6 months); no recent substantial weight change (>20 lbs in past 6 months); and no recent flu shot or other vaccination (<1 month). We excluded or deferred the listed conditions above for their known substantial effects on systemic metabolism and/or their temporary effects on the gut microbiota directly (endoscopy and antibiotics) [123,124].

Eligible participants received a study packet in the mail, including the consent form to sign and return (and a copy to keep), a stool collection kit, instructions for stool collection, and a Stool Collection Questionnaire. The stool collection kit consisted of a stool collection container (white tub with lid; Fisher), a toilet adaptor tray, a sealed sample collection tube (smaller, inner tube; Sarstedt) containing 10 glass beads (3mm; Fisher) and 5ml RNAlater® (Ambion), an outer, protective tube with screw cap (Sarstedt), a pair of exam gloves, a biohazard specimen bag, and an absorbent pad. The packet also included a return mailer box with prepaid postage.

The collection protocol took less than 15 minutes for most participants. Participants were asked to collect the sample between Sunday and Wednesday and were provided with easy-to-follow instructions with photos for collecting the sample: insert the white collection container in the middle of the toilet seat adaptor and place the adaptor underneath the toilet seat, collect stool into the container, put on the gloves and open the smaller tube with clear liquid in it, put a small amount (2 pea-sized scoops) of stool into the small tube using the scoop that is attached to the lid, screw the lid tightly so that the liquid inside does not leak, and shake the tube vigorously for 30 seconds so that the stool dissolves in the liquid. Participants were then asked to put the smaller tube inside the larger tube, seal the tubes inside the plastic biohazard bag with the absorbent pad, and store the bag in their home freezer overnight before mailing it the next morning along with the signed consent form and

the completed questionnaire. Stool samples were stored at -80°C upon receipt at study centers until ready for processing. The Stool Collection Questionnaire inquired on the sampling details (date, time, overnight freezing, any problems), overall health (body weight, health concerns), past year history of antibiotic or antifungal medication use, gastric procedures, probiotic pill or laxative use, rural or urban childhood (birth to age 3) environment, and past year consumption of probiotic foods, special diets or artificial sweeteners. Participants received a thank-you letter and a gift certificate or a check in the mail for their time and effort.

Between February 2013 and June 2015, the participation rate was 49.7% with some regional difference between California and Hawaii likely due to greater residential mobility among mainland participants (Table 1). The participation rate, with the small amount of compensation, was similar to that for other sample collections conducted in the MEC. Participation rate was also similar between men and women and highest among Whites (60.9%), followed by Native Hawaiians (53.6%), Japanese Americans (48.7%), African Americans (47.9%) and Latinos (45.0%).

Compared to MEC participants who provided a stool sample, those who refused or did not respond to mailed and telephone requests were less likely to be White and were younger at baseline by an average of 2.6 years. After adjustment for sex, ethnicity and age, the two groups were similar on marital status, number of children, history of hormone therapy use among women, intakes of energy and macronutrients, and diagnosis of diabetes. However, those who did not provide a sample had an average of 0.4 fewer years of education, spent less time in moderate and vigorous activity, had a higher body mass index by an average of 0.5 kg/m^2 , and were 5% more likely to have been smokers or to have a diagnosis of hypertension.

For processing, stool samples are shipped on dry ice from study centers in Honolulu and Los Angeles to Seattle. Among the samples collected to date, we found that, on average the amount of time it took for samples to be shipped to the study center was 2.9 days and that 70.6% of samples were received in 3 days or less (Table 2). Upon receipt in Seattle, DNA is extracted following a protocol optimized for stool [27]. Based on the protocol in which participants are instructed to add 2 pea-size scoops of stool into 5ml RNeasy® Lysis Buffer, we obtain on average $57.68\text{ }\mu\text{g DNA/ml}$. Of the samples extracted for DNA to date, only 6.5% have failed to provide sufficient bacterial DNA for sequencing. The V1-V3 region of the 16S rRNA gene is sequenced using the Illumina MiSeq platform to obtain 2x300 bp paired-end reads. Joining of forward and reverse reads, quality filtering, and analysis are performed using QIIME [74]. Analysis of these data will further elucidate the contribution of specimen collection and handling on characterization of the gut microbial community within the MEC.

CONCLUSION

There are many steps in the collection, processing, and analysis of gut microbiome samples that can influence the data generated and potentially the interpretation of a study. Although the studies described in this review typically show that methodological variation tends to be less than inter-individual variation in the gut microbiome, technical issues may become a

bigger concern as the field focuses more on specific differences (especially for taxa that are less abundant and at lower taxonomic levels), and addresses functional shifts using metagenomic sequencing and microbial multi-omics approaches. Continued investigation with larger studies, such as the Microbiome Quality Control project (MBQC) [71], will allow for the efficient assessment of multiple sources of variation along with associated interactions. As epidemiologists develop larger and more complex microbiome studies with these issues in mind, standardized methods, such as those developed by the HMP [24], need to be adopted and built upon to ensure data quality and better comparability among studies.

Acknowledgments

This work was supported in part by grants from the National Institutes of Health (P01 CA168530, T32 CA009168, U01 CA164973)

REFERENCES CITED

1. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, et al. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012; 486:207–214. [PubMed: 22699609]
2. Li J, Jia H, Cai X, Zhong H, Feng Q, et al. An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol*. 2014; 32:834–841. [PubMed: 24997786]
3. Wang WL, Xu SY, Ren ZG, Tao L, Jiang JW, et al. Application of metagenomics in the human gut microbiome. *World J Gastroenterol*. 2015; 21:803–814. [PubMed: 25624713]
4. Gillevet P, Sikaroodi M, Keshavarzian A, Mutlu EA. Quantitative assessment of the human gut microbiome using multitag pyrosequencing. *Chem Biodivers*. 2010; 7:1065–1075. [PubMed: 20491064]
5. Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben-Amor K, Akkermans AD, et al. Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl Environ Microbiol*. 2002; 68:3401–3407. [PubMed: 12089021]
6. Durban A, Abellan JJ, Jimenez-Hernandez N, Ponce M, Ponce J, et al. Assessing gut microbial diversity from feces and rectal mucosa. *Microb Ecol*. 2011; 61:123–133. [PubMed: 20734040]
7. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, et al. Diversity of the human intestinal microbial flora. *Science*. 2005; 308:1635–1638. [PubMed: 15831718]
8. Bajaj JS, Hylemon PB, Ridlon JM, Heuman DM, Daita K, et al. Colonic mucosal microbiome differs from stool microbiome in cirrhosis and hepatic encephalopathy and is linked to cognition and inflammation. *Am J Physiol Gastrointest Liver Physiol*. 2012; 303:G675–G685. [PubMed: 22821944]
9. Momozawa Y, Deffontaine V, Louis E, Medrano JF. Characterization of bacteria in biopsies of colon and stools by high throughput sequencing of the V2 region of bacterial 16S rRNA gene in human. *PLoS One*. 2011; 6:e16952. [PubMed: 21347324]
10. Zilberstein B, Quintanilha AG, Santos MA, Pajek D, Moura EG, et al. Digestive tract microbiota in healthy volunteers. *Clinics (Sao Paulo)*. 2007; 62:47–54. [PubMed: 17334549]
11. Zoetendal EG, Raes J, van den Bogert B, Arumugam M, Boonink CCGM, et al. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *ISME J*. 2012; 6:1415–1426. [PubMed: 22258098]
12. Araujo-Perez F, McCoy AN, Okechukwu C, Carroll IM, Smith KM, et al. Differences in microbial signatures between rectal mucosal biopsies and rectal swabs. *Gut Microbes*. 2012; 3:530–535. [PubMed: 23060016]
13. Tedjo DI, Jonkers DM, Savelkoul PH, Masclee AA, van Best N, et al. The effect of sampling and storage on the fecal microbiota composition in healthy and diseased subjects. *PLoS One*. 2015; 10:e0126685. [PubMed: 26024217]

14. Huse SM, Young VB, Morrison HG, Antonopoulos DA, Kwon J, et al. Comparison of brush and biopsy sampling methods of the ileal pouch for assessment of mucosa-associated microbiota of human subjects. *Microbiome*. 2014; 2:5. [PubMed: 24529162]
15. Dominianni C, Wu J, Hayes RB, Ahn J. Comparison of methods for fecal microbiome biospecimen collection. *BMC Microbiol*. 2014; 14:103. [PubMed: 24758293]
16. Sinha R, Chen J, Amir A, Vogtmann E, Shi J, et al. Collecting Fecal Samples for Microbiome Analyses in Epidemiology Studies. *Cancer Epidemiol Biomarkers Prev*. 2015
17. Imrit K, Goldfischer M, Wang J, Green J, Levine J, et al. Identification of bacteria in formalin-fixed, paraffin-embedded heart valve tissue via 16S rRNA gene nucleotide sequencing. *J Clin Microbiol*. 2006; 44:2609–2611. [PubMed: 16825394]
18. Mima K, Nishihara R, Qian ZR, Cao Y, Sukawa Y, et al. *Fusobacterium nucleatum* in colorectal carcinoma tissue and patient prognosis. *Gut*. 2015
19. Xuan C, Shamonki JM, Chung A, Dinome ML, Chung M, et al. Microbial dysbiosis is associated with human breast cancer. *PLoS One*. 2014; 9:e83744. [PubMed: 24421902]
20. Gilbert MT, Haselkorn T, Bunce M, Sanchez JJ, Lucas SB, et al. The isolation of nucleic acids from fixed, paraffin-embedded tissues-which methods are useful when? *PLoS One*. 2007; 2:e537. [PubMed: 17579711]
21. Farrugia A, Keyser C, Ludes B. Efficiency evaluation of a DNA extraction and purification protocol on archival formalin-fixed and paraffin-embedded tissue. *Forensic Sci Int*. 2010; 194:e25–28. [PubMed: 19781880]
22. Jones TF, Bulens SN, Gettner S, Garman RL, Vugia DJ, et al. Use of stool collection kits delivered to patients can improve confirmation of etiology in foodborne disease outbreaks. *Clin Infect Dis*. 2004; 39:1454–1459. [PubMed: 15546081]
23. Itzkowitz SH, Jandorf L, Brand R, Rabeneck L, Schroy PC 3rd, et al. Improved fecal DNA test for colorectal cancer screening. *Clin Gastroenterol Hepatol*. 2007; 5:111–117. [PubMed: 17161655]
24. Human Microbiome Project C. A framework for human microbiome research. *Nature*. 2012; 486:215–221. [PubMed: 22699610]
25. Walker AW, Duncan SH, McWilliam Leitch EC, Child MW, Flint HJ. pH and peptide supply can radically alter bacterial populations and short-chain fatty acid ratios within microbial communities from the human colon. *Appl Environ Microbiol*. 2005; 71:3692–3700. [PubMed: 16000778]
26. Albenberg L, Esipova TV, Judge CP, Bittinger K, Chen J, et al. Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota. *Gastroenterology*. 2014; 147:1055–1063. e1058. [PubMed: 25046162]
27. Li F, Hullar MA, Lampe JW. Optimization of terminal restriction fragment polymorphism (TRFLP) analysis of human gut microbiota. *J Microbiol Methods*. 2007; 68:303–311. [PubMed: 17069911]
28. Wesolowska-Andersen A, Bahl MI, Carvalho V, Kristiansen K, Sicheritz-Ponten T, et al. Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome*. 2014; 2:19. [PubMed: 24949196]
29. Gorzelak MA, Gill SK, Tasnim N, Ahmadi-Vand Z, Jay M, et al. Methods for Improving Human Gut Microbiome Data by Reducing Variability through Sample Processing and Storage of Stool. *PLoS One*. 2015; 10:e0134802. [PubMed: 26252519]
30. Wu GD, Lewis JD, Hoffmann C, Chen YY, Knight R, et al. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC Microbiol*. 2010; 10:206. [PubMed: 20673359]
31. Hang J, Desai V, Zavaljevski N, Yang Y, Lin X, et al. 16S rRNA gene pyrosequencing of reference and clinical samples and investigation of the temperature stability of microbiome profiles. *Microbiome*. 2014; 2:31. [PubMed: 25228989]
32. Roesch LF, Casella G, Simell O, Krischer J, Wasserfall CH, et al. Influence of fecal sample storage on bacterial community diversity. *Open Microbiol J*. 2009; 3:40–46. [PubMed: 19440250]
33. Cardona S, Eck A, Cassellas M, Gallart M, Alastrue C, et al. Storage conditions of intestinal microbiota matter in metagenomic analysis. *BMC Microbiol*. 2012; 12:158. [PubMed: 22846661]

34. Rubin BE, Gibbons SM, Kennedy S, Hampton-Marcell J, Owens S, et al. Investigating the impact of storage conditions on microbial community composition in soil samples. *PLoS One*. 2013; 8:e70460. [PubMed: 23936206]
35. Fouhy F, Deane J, Rea MC, O'Sullivan O, Ross RP, et al. The effects of freezing on faecal microbiota as determined using MiSeq sequencing and culture-based investigations. *PLoS One*. 2015; 10:e0119355. [PubMed: 25748176]
36. Flores R, Shi J, Yu G, Ma B, Ravel J, et al. Collection media and delayed freezing effects on microbial composition of human stool. *Microbiome*. 2015; 3:33. [PubMed: 26269741]
37. Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, et al. Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A*. 2014; 111:E2329–2338. [PubMed: 24843156]
38. Voigt AY, Costea PI, Kultima JR, Li SS, Zeller G, et al. Temporal and technical variability of human gut metagenomes. *Genome Biol*. 2015; 16:73. [PubMed: 25888008]
39. Vlckova K, Mrazek J, Kopečný J, Petrzeková KJ. Evaluation of different storage methods to characterize the fecal bacterial communities of captive western lowland gorillas (*Gorilla gorilla gorilla*). *J Microbiol Methods*. 2012; 91:45–51. [PubMed: 22828127]
40. McCarthy A, Chiang E, Schmidt ML, Deneff VJ. RNA preservation agents and nucleic acid extraction method bias perceived bacterial community composition. *PLoS One*. 2015; 10:e0121659. [PubMed: 25798612]
41. Choo JM, Leong LE, Rogers GB. Sample storage conditions significantly influence faecal microbiome profiles. *Sci Rep*. 2015; 5:16350. [PubMed: 26572876]
42. 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; 307:80–86. [PubMed: 20412303]
43. 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:e46953. [PubMed: 23071673]
44. Kennedy NA, Walker AW, Berry SH, Duncan SH, Farquarson FM, et al. The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS One*. 2014; 9:e88982. [PubMed: 24586470]
45. Larsen AM, Mohammed HH, Arias CR. Comparison of DNA extraction protocols for the analysis of gut microbiota in fishes. *FEMS Microbiol Lett*. 2015:362.
46. Reck M, Tomasch J, Deng ZL, Jarek M, Husemann P, et al. Stool metatranscriptomics: A technical guideline for mRNA stabilisation and isolation. *BMC Genomics*. 2015:16. [PubMed: 25613955]
47. Zoetendal EG, Booiijink CC, Klaassens ES, Heilig HG, Kleerebezem M, et al. Isolation of RNA from bacterial samples of the human gastrointestinal tract. *Nat Protocols*. 2006; 1:954–959. [PubMed: 17406329]
48. Santiago A, Panda S, Mengels G, Martinez X, Azpiroz F, et al. Processing faecal samples: a step forward for standards in microbial community analysis. *BMC Microbiol*. 2014; 14:112. [PubMed: 24884524]
49. Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ, et al. 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. *Microbiome*. 2015; 3:26. [PubMed: 26120470]
50. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One*. 2012; 7:e33865. [PubMed: 22457796]
51. Satinsky BM, Gifford SM, Crump BC, Moran MA. Use of internal standards for quantitative metatranscriptome and metagenome analysis. *Methods Enzymol*. 2013; 531:237–250. [PubMed: 24060124]
52. Stewart FJ, Ottesen EA, DeLong EF. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J*. 2010; 4:896–907. [PubMed: 20220791]

53. Gilbert JA, Bao Y, Wang H, Sansone SA, Edmunds SC, et al. Report of the 13(th) Genomic Standards Consortium Meeting, Shenzhen, China, March 4–7, 2012. *Stand Genomic Sci.* 2012; 6:276–286. [PubMed: 22768370]
54. Field D, Sterk P, Kottmann R, De Smet JW, Amaral-Zettler L, et al. Genomic standards consortium projects. *Stand Genomic Sci.* 2014; 9:599–601. [PubMed: 25197446]
55. Reuter JA, Spacek DV, Snyder MP. High-throughput sequencing technologies. *Mol Cell.* 2015; 58:586–597. [PubMed: 26000844]
56. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One.* 2011; 6:e27310. [PubMed: 22194782]
57. Jumpstart Consortium Human Microbiome Project Data Generation Working G. Evaluation of 16S rDNA-based community profiling for human microbiome research. *PLoS One.* 2012; 7:e39315. [PubMed: 22720093]
58. Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods.* 2007; 69:330–339. [PubMed: 17391789]
59. Sharpton TJ. An introduction to the analysis of shotgun metagenomic data. *Front Plant Sci.* 2014; 5:209. [PubMed: 24982662]
60. Morgan XC, Huttenhower C. Meta’omic analytic techniques for studying the intestinal microbiome. *Gastroenterology.* 2014; 146:1437–1448. e1431. [PubMed: 24486053]
61. Pinto AJ, Raskin L. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS One.* 2012; 7:e43093. [PubMed: 22905208]
62. Ahn JH, Kim BY, Song J, Weon HY. Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. *J Microbiol.* 2012; 50:1071–1074. [PubMed: 23274999]
63. Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol.* 2005; 71:8966–8969. [PubMed: 16332901]
64. McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol.* 2014; 10:e1003531. [PubMed: 24699258]
65. Farris MH, Olson JB. Detection of Actinobacteria cultivated from environmental samples reveals bias in universal primers. *Lett Appl Microbiol.* 2007; 45:376–381. [PubMed: 17897379]
66. Lee ZM, Bussema C 3rd, Schmidt TM. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Res.* 2009; 37:D489–493. [PubMed: 18948294]
67. Kembel SW, Wu M, Eisen JA, Green JL. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Comput Biol.* 2012; 8:e1002743. [PubMed: 23133348]
68. 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:e30087. [PubMed: 22347999]
69. Schirmer M, Ijaz UZ, D’Amore R, Hall N, Sloan WT, et al. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* 2015; 43:e37. [PubMed: 25586220]
70. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, et al. Moving pictures of the human microbiome. *Genome Biol.* 2011; 12:R50. [PubMed: 21624126]
71. 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. [PubMed: 26653756]
72. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods.* 2013; 10:57–59. [PubMed: 23202435]
73. Zhou Q, Su X, Ning K. Assessment of quality control approaches for metagenomic data analysis. *Sci Rep.* 2014; 4:6957. [PubMed: 25376098]
74. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010; 7:335–336. [PubMed: 20383131]

75. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. Introducing mothur: Open-Source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 2009; 75:7537–7541. [PubMed: 19801464]
76. Hugenholtz P, Goebel BM, Pace NR. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol.* 1998; 180:4765–4774. [PubMed: 9733676]
77. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, et al. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Res.* 2014; 42:D643–648. [PubMed: 24293649]
78. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol.* 2006; 72:5069–5072. [PubMed: 16820507]
79. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 2009; 37:D141–145. [PubMed: 19004872]
80. Schloss PD, Handelsman J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol.* 2005; 71:1501–1506. [PubMed: 15746353]
81. Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J. Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PLoS One.* 2014; 9:e94249. [PubMed: 24722003]
82. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. *PLoS Comput Biol.* 2012; 8:e1002687. [PubMed: 23028285]
83. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, et al. Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol.* 2012; 8:e1002606. [PubMed: 22807668]
84. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. *Nat Methods.* 2013; 10:1200–1202. [PubMed: 24076764]
85. Costea PI, Zeller G, Sunagawa S, Bork P. A fair comparison. *Nat Methods.* 2014; 11:359–359. [PubMed: 24681719]
86. Weiss SJ, Xu Z, Amir A, Peddada S, Bittinger K, et al. Effects of library size variance, sparsity, and compositionality on the analysis of microbiome data. *PeerJ PrePrints.* 2015:2167–9843.
87. Li H. Microbiome, metagenomics, and high-dimensional compositional data analysis. *Annu Rev Stat Appl.* 2015; 2:73–94.
88. Lovell D, Pawlowsky-Glahn V, Egozcue JJ, Marguerat S, Bahler J. Proportionality: a valid alternative to correlation for relative data. *PLoS Comput Biol.* 2015; 11:e1004075. [PubMed: 25775355]
89. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* 2011; 12doi: 10.1186/Gb-2011-1112-1186-R1160
90. Xu L, Paterson AD, Turpin W, Xu W. Assessment and Selection of Competing Models for Zero-Inflated Microbiome Data. *PLoS One.* 2015; 10:e0129606. [PubMed: 26148172]
91. Legendre, P.; Legendre, LF. Numerical ecology. Elsevier; 2012.
92. Zuur, A.; Ieno, EN.; Smith, GM. Analysing ecological data. Springer Science & Business Media; 2007.
93. Kuczynski J, Lauber CL, Walters WA, Parfrey LW, Clemente JC, et al. Experimental and analytical tools for studying the human microbiome. *Nat Rev Genet.* 2012; 13:47–58. [PubMed: 22179717]
94. Di Bella JM, Bao Y, Gloor GB, Burton JP, Reid G. High throughput sequencing methods and analysis for microbiome research. *J Microbiol Methods.* 2013; 95:401–414. [PubMed: 24029734]
95. Lozupone CA, Knight R. Species divergence and the measurement of microbial diversity. *FEMS Microbiol Rev.* 2008; 32:557–578. [PubMed: 18435746]
96. Chen J, Bittinger K, Charlson ES, Hoffmann C, Lewis J, et al. Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics.* 2012; 28:2106–2113. [PubMed: 22711789]

97. Matsen, FAt; Evans, SN. Edge principal components and squash clustering: using the special structure of phylogenetic placement data for sample comparison. *PLoS One*. 2013; 8:e56859. [PubMed: 23505415]
98. Purdom E. Analysis of data matrix and a graph: metagenomic data and the phylogenetic tree. *Ann Appl Stat*. 2011; 5:813–820.
99. Fukuyama J, McMurdie PJ, Dethlefsen L, Relman DA, Holmes S. Comparisons of distance methods for combining covariates and abundances in microbiome studies. *Pac Symp Biocomput*. 2012:213–224. [PubMed: 22174277]
100. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*. 2001; 26:32–46.
101. Zhao N, Chen J, Carroll IM, Ringel-Kulka T, Epstein MP, et al. Testing in Microbiome-Profiling Studies with MiRKAT, the Microbiome Regression-Based Kernel Association Test. *Am J Hum Genet*. 2015; 96:797–807. [PubMed: 25957468]
102. Buhlmann P, Kalisch M, Meier L. High-dimensional statistics with a view toward applications in Biology. *Annu Rev Stat Appl*. 2014; 1:255–278.
103. Randolph, TW.; Zhao, S.; Copeland, WK.; Hullar, M.; Shojaiie, A. Kernel-Penalized regression for analysis of microbiome data. 2015. arXiv: <http://arxiv.org/abs/1511.00297>
104. Feigelson HS, Bischoff K, Ardini MA, Ravel J, Gail MH, et al. Feasibility of self-collection of fecal specimens by randomly sampled women for health-related studies of the gut microbiome. *BMC Res Notes*. 2014; 7:204. [PubMed: 24690120]
105. Schultze A, Akmatov MK, Andrzejak M, Karras N, Kemmling Y, et al. Comparison of stool collection on site versus at home in a population-based study : feasibility and participants' preference in Pretest 2 of the German National Cohort. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2014; 57:1264–1269. [PubMed: 25293889]
106. de Wit MA, Koopmans MP, Kortbeek LM, Wannet WJ, Vinje J, et al. Sensor, a population-based cohort study on gastroenteritis in the Netherlands: incidence and etiology. *Am J Epidemiol*. 2001; 154:666–674. [PubMed: 11581101]
107. Wheeler JG, Sethi D, Cowden JM, Wall PG, Rodrigues LC, et al. Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. The Infectious Intestinal Disease Study Executive. *BMJ*. 1999; 318:1046–1050. [PubMed: 10205103]
108. Powell EO. Growth rate and generation time of bacteria, with special reference to continuous culture. *J Gen Microbiol*. 1956; 15:492–511. [PubMed: 13385433]
109. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, et al. Bacterial community variation in human body habitats across space and time. *Science*. 2009; 326:1694–1697. [PubMed: 19892944]
110. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, et al. The long-term stability of the human gut microbiota. *Science*. 2013; 341:1237439. [PubMed: 23828941]
111. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, et al. Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A*. 2011; 108(Suppl 1):4578–4585. [PubMed: 20668239]
112. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012; 486:222–227. [PubMed: 22699611]
113. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci USA*. 2010; 107:14691–14696. [PubMed: 20679230]
114. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science*. 2011; 334:105–108. [PubMed: 21885731]
115. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. 2014; 505:559–563. [PubMed: 24336217]
116. David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, et al. Host lifestyle affects human microbiota on daily timescales. *Genome Biol*. 2014; 15:R89. [PubMed: 25146375]

117. Perez-Cobas AE, Gosalbes MJ, Friedrichs A, Knecht H, Artacho A, et al. Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. *Gut*. 2013; 62:1591–1601. [PubMed: 23236009]
118. Jernberg C, Lofmark S, Edlund C, Jansson JK. Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology*. 2010; 156:3216–3223. [PubMed: 20705661]
119. 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):4554–4561. [PubMed: 20847294]
120. Duncan SH, Lopley GE, Holtrop G, Ince J, Johnstone AM, et al. Human colonic microbiota associated with diet, obesity and weight loss. *Int J Obes*. 2008; 32:1720–1724.
121. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. *Nature*. 2006; 444:1022–1023. [PubMed: 17183309]
122. Frankenfeld CL, Atkinson C, Thomas WK, Gonzalez A, Jokela T, et al. High concordance of daidzein-metabolizing phenotypes in individuals measured 1 to 3 years apart. *Br J Nutr*. 2005; 94:873–876. [PubMed: 16351761]
123. Jakobsson HE, Jernberg C, Andersson AF, Sjolund-Karlsson M, Jansson JK, et al. Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. *PLoS One*. 2010; 5:e9836. [PubMed: 20352091]
124. O'Brien CL, Allison GE, Grimpén F, Pavli P. Impact of colonoscopy bowel preparation on intestinal microbiota. *PLoS One*. 2013; 8:e62815. [PubMed: 23650530]

Table 1

Recruitment and participation in stool sample collection in the MEC

	Hawaii	California	TOTAL
Contacted	5255	4965	10220
Could not be reached	365 (6.9%)	908 (18.3%)	1273 (12.5%)
Refused	1561 (29.7%)	1315 (26.5%)	2876 (28.1%)
Deceased or too ill	510 (9.7%)	784 (15.8%)	1294 (12.7%)
Ineligible	477 (9.1%)	198 (4.0%)	675 (6.6%)
Participated	2342 (44.6%)	1760 (35.4%)	4102 (40.1%)
Participated (among eligible and alive)	54.9%	44.2%	49.7%

Table 2

Distribution of days from stool sample collection to receipt at study center at the two MEC sites and DNA yield from stool samples as collected by participants per protocol.

	Hawaii	California	Total
Days from collection to shipment			
<1	91 (4.6%)	250 (30.9%)	341 (12.3%)
1–2	512 (26.2%)	230 (28.4%)	742 (26.9%)
2–3	737 (37.7%)	130 (16.1%)	867 (31.4%)
3+	615 (31.5%)	199 (24.6%)	814 (29.4%)
DNA yield (µg/ml)	54.17 (35.3)	66.19 (48.5)	57.68 (40.0)

Displayed as n(%) or mean(SD)