JCI The Journal of Clinical Investigation

Next-generation sequencing: insights to advance clinical investigations of the microbiome

Caroline R. Wensel, … , Steven L. Salzberg, Cynthia L. Sears

J Clin Invest. 2022;132(7):e154944. https://doi.org/10.1172/JCI154944.

Review Series

Next-generation sequencing (NGS) technology has advanced our understanding of the human microbiome by allowing for the discovery and characterization of unculturable microbes with prediction of their function. Key NGS methods include 16S rRNA gene sequencing, shotgun metagenomic sequencing, and RNA sequencing. The choice of which NGS methodology to pursue for a given purpose is often unclear for clinicians and researchers. In this Review, we describe the fundamentals of NGS, with a focus on 16S rRNA and shotgun metagenomic sequencing. We also discuss pros and cons of each methodology as well as important concepts in data variability, study design, and clinical metadata collection. We further present examples of how NGS studies of the human microbiome have advanced our understanding of human disease pathophysiology across diverse clinical contexts, including the development of diagnostics and therapeutics. Finally, we share insights as to how NGS might further be integrated into and advance microbiome research and clinical care in the coming years.

Next-generation sequencing: insights to advance clinical investigations of the microbiome

Caroline R. Wensel,[1] Jennifer L. Pluznick,[2] Steven L. Salzberg,[3,4,5] and Cynthia L. Sears[1,6]

[1]Department of Medicine and [2]Department of Physiology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [3]Department of Biomedical Engineering, [4]Department of Computer Science, and [5]Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland, USA. [6]Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

Next-generation sequencing (NGS) technology has advanced our understanding of the human microbiome by allowing for the discovery and characterization of unculturable microbes with prediction of their function. Key NGS methods include 16S rRNA gene sequencing, shotgun metagenomic sequencing, and RNA sequencing. The choice of which NGS methodology to pursue for a given purpose is often unclear for clinicians and researchers. In this Review, we describe the fundamentals of NGS, with a focus on 16S rRNA and shotgun metagenomic sequencing. We also discuss pros and cons of each methodology as well as important concepts in data variability, study design, and clinical metadata collection. We further present examples of how NGS studies of the human microbiome have advanced our understanding of human disease pathophysiology across diverse clinical contexts, including the development of diagnostics and therapeutics. Finally, we share insights as to how NGS might further be integrated into and advance microbiome research and clinical care in the coming years.

Introduction

The number of microbial cells that reside on and in us rivals the number of our own cells (1). In health, we, the host, and microbes live in symbiosis. However, many illnesses are defined by or associated with microbial dysbiosis. These include both communicable diseases such as tuberculosis (2) and syphilis (3), and non-communicable diseases like inflammatory bowel disease (4), diabetes (5), obesity (6), and cancer (7). We have known since the time of Koch and the discovery of *Mycobacterium tuberculosis* that our microbial inhabitants affect our health status. However, how we characterize these organisms has drastically changed (8–10). Koch's postulates laid a framework for assessing microbial causes of disease through culturing methods. Indeed, Koch's postulates are still relevant, more than 130 years after they were first published; however, we now also recognize the importance and vast diversity of unculturable microbes (11). Advances in technology like next-generation sequencing (NGS) have led to an explosion in the discovery and characterization of microbes, because NGS methods do not rely on traditional culture techniques and can thus detect the unculturable microbes (Table 1). In fact, complementing of traditional culture methods with NGS has already been implemented in many clinical microbiology laboratories because of its potential to address severe, insidious infections (12). Advantages of NGS include its ability to identify more unique species than traditional culture methods, and the capacity to perform parallel sequencing of multiple samples, which, with the earlier, low-throughput Sanger sequencing technology, was not technically feasible.

The purpose of this Review is to discuss crucial features relating to NGS in translational research and clinical care. We first discuss the fundamentals of NGS and compare common methodologies as well as sources of data variability and important study design considerations. Then, we present select examples of how NGS has altered our collective understanding of disease pathogenesis. Finally, we offer insights as to how NGS might further be integrated into and advance clinical care in the coming years with the aim of helping researchers and clinicians consider the impact of NGS on disease diagnostics and therapeutics.

Fundamental considerations in the use of NGS

What are the fundamentals of NGS? An initial question in studying the gut microbiota is which microbes are present in a given sample. Subsequent inquiries, addressable by NGS analyses, include determining the relative abundance and predictive functional profiles of the microbes present, as well as understanding intraspecies and population heterogeneity (13). NGS methods address these questions by directly sequencing microbial DNA or RNA, for example, in fecal, blood, and/or tissue samples. With the improving affordability of NGS, the two primary NGS methodologies now in use are amplicon sequencing and shotgun metagenomic sequencing; however, RNA sequencing is also a valid and, in some ways, superior method for microbial characterization, as it allows for determination of the transcriptome, representing a further step to define microbiota function (14, 15).

One of the most common NGS methods for bacterial identification and characterization is amplicon sequencing. Amplicon sequencing involves first amplifying a region of the DNA via PCR, and then sequencing the resultant product. The target for PCR amplification is, most commonly, the bacterial 16S ribosomal RNA (rRNA) gene (Figure 1). For this reason, amplicon sequencing is

Conflict of interest: CLS has received research funding to her institution from Janssen and Bristol Myers Squibb and has served on a Ferring Pharmaceuticals clinical advisory board and as a reviewer for UpToDate.

Reference information: *J Clin Invest*. 2022;132(7):e154944.
https://doi.org/10.1172/JCI154944.

Table 1. Proposed Koch's postulates for NGS

Original[A]	Molecular[B]	Next-generation sequencing[C]
• The microorganism is found in abundance in diseased but not in healthy individuals	• The virulence gene is found in pathogenic but not nonpathogenic microbial strains	• Individual microorganisms or communities of microorganisms, as identified by sequencing, differ in abundance, organization, and/or function in diseased vs. healthy individuals
• The microorganism is able to be isolated from the diseased host and grown in pure culture	• Deletion or inactivation of the virulence gene leads to loss of microbe pathogenicity	• Community virulence or functional consequences may or may not depend on specific, well-defined microbe virulence genes
• The cultured microorganism causes disease when inoculated into a healthy host	• Reactivation or allelic replacement of the gene restores microbe pathogenicity	• Community modification and/or elimination of specific community members alleviates the disease state
• The microorganism can be re-isolated from the inoculated host and is the same as the original microorganism		
• Elimination of the microbe from the host alleviates disease		

[A]The first four postulates are derived directly from Koch's original postulates, whereas the fifth is derived from Evans (138). Koch's original postulates do not account for viruses, parasites, unculturable bacteria, and/or the concept of host colonization with a potential pathogen. [B]Focuses on genes that make a microbe virulent (10, 139). [C]Focuses on use of nucleic acid sequences rather than culture or entire genes to find emerging pathogens (8, 9).

also referred to as 16S rRNA sequencing or analysis. The use of the 16S rRNA gene to characterize uncultured microbes was first described by Lane et al. in 1985 (16). The 16S rRNA gene is an ideal target because it is highly conserved and ubiquitous among bacteria (without it, bacteria would be unable to translate mRNA into proteins and thus be nonfunctional) and it also contains nine hypervariable regions (V1–V9) that differ between bacterial species and genera (Figure 1). Thus, PCR primers can be designed such that forward and reverse primers bind to conserved regions but amplify an intervening variable region. Typically only a subset of the variable regions are targeted for sequencing in a given study (e.g., V1–V3, V4–V5) to limit the amount and, thus, time and cost of sequencing. However, it is important to note that no one region adequately differentiates all bacteria (17), and sequencing of select hypervariable regions can yield differing data interpretation (17–19). For example, amplification of certain hypervariable regions may bias results, leading to under- or overrepresentation of taxa (18), but may also be advantageous for distinguishing between certain species within a genus (17). Recently, NGS sequencing of the full 16S rRNA gene has emerged and, using increasingly sophisticated analytical methods, may provide both species and strain resolution in microbiota communities (20).

After PCR amplification of the selected hypervariable regions, the resulting amplicons are sequenced, followed by data "cleaning." Data cleaning involves multiple steps, such as adapter and primer sequence trimming, removal of low-quality bases and sequences from reads, and removal of sequences matching a control library such as the PhiX Control (Illumina), chimeric sequences, and human contaminant reads, as well as chloroplast and mitochondrial contaminants. Subsequent analyses lead to organization of the sequence data into, most often, operational taxonomic units (OTUs). OTUs are distance-based clusters of sequences, initially constructed without a reference database (21). An OTU sequence identity greater than 97% (or with up to 3% dissimilarity) is typically estimated to define a species, while OTUs with sequence similarities of 95% and 80% are used to define genus and phylum, respectively (21). Taxonomic identification is then inferred by computational alignment to reference 16S rRNA sequence databases such as the Ribosomal Database Project (RDP) (22), SILVA (23), or Greengenes (24). OTUs and identified taxa are then used for downstream analysis. An alternative, less frequently used nondistance-based analytical approach for amplicon sequencing relies on exact nucleotide matching to yield amplicon sequence variants (ASVs). ASV taxon assignments are dependent on the quality of reference databases (25). Additionally, ASVs have the potential to split single genomes into multiple clusters, because most bacterial cells possess more than one rRNA gene copy and these, not infrequently, differ in nucleotide sequence (26). While each method (OTUs versus ASVs) has proponents (26, 27), importantly, both are computational approaches to estimate taxonomy. For unculturable microbes, NGS data alone produce "candidate species," whereas firmer classification of cultured bacterial species is possible using both phenotypic and genome sequence data (28).

In contrast to amplicon sequencing, shotgun metagenomic sequencing and RNA sequencing analyze all the DNA or RNA in a given sample, respectively. For shotgun metagenomic sequencing, after extraction, the DNA is randomly fragmented, and barcodes and adapters are ligated to the ends of each segment to facilitate sample identification and DNA sequencing. The resultant reads are cleaned and subsequently aligned to a reference database to identify taxa and functional potential. The primary reference databases are usually Reference Sequence (RefSeq; ref. 29) and GenBank (30). These are large databases containing all publicly available genomes. Smaller pathogen-focused databases such as Pathosystems Resource Integration Center (PATRIC; ref. 31) and the Eukaryotic Pathogen Database (EuPathDB; ref. 32) are also used. The RNA sequencing workflow is similar to that for shotgun metagenomic sequencing; however, after fragmentation, the RNA segments are reverse transcribed, using PCR, into complementary DNA (cDNA), which is then processed using the DNA sequencing pipeline. Figure 2 provides an overview of NGS processes. Because of their diverse methodologies, 16S rRNA amplicon, shotgun metagenomic, and RNA sequencing each have advantages and drawbacks. These are discussed below. Choosing

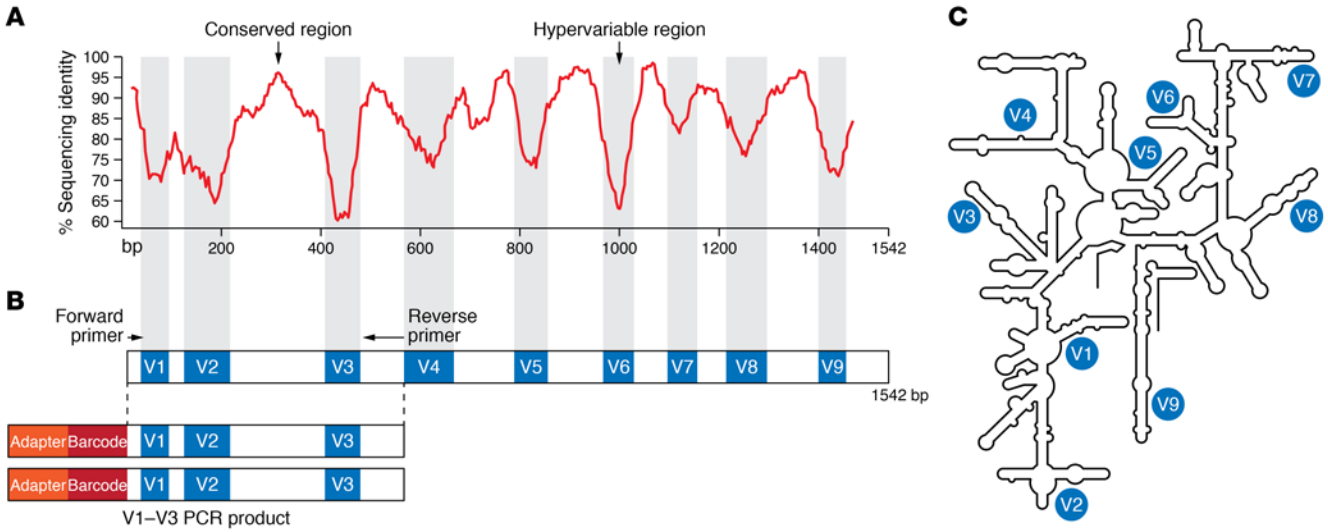

Figure 1. Bacterial 16S rRNA gene. (**A**) Percentage sequence identity of conserved and hypervariable regions of the bacterial 16S rRNA gene. Adapted with permission from the *Journal of Microbiological Methods* (17) and Ilona Lehtinen (137). (**B**) Illustration of conserved and hypervariable regions corresponding to **A** and PCR amplification of the V1–V3 region of the bacterial 16S rRNA gene. Adapted with permission from Humana Press (148). (**C**) Schematic of 16S rRNA gene structure with hypervariable regions (V1–V9) labeled.

one method over the others requires comparison and consideration of study goals. Several recent reviews and books have provided guides to microbiome analysis (21, 33).

Direct comparisons between NGS methods. Although comparisons of 16S rRNA sequencing and shotgun metagenomics exist for a variety of samples, including those from humans (14, 15, 34–40), laboratory model organisms (13, 39), plants (39, 41), soil (42), and water (43), overall, direct method comparisons for human samples are limited. Comparisons with RNA sequencing and across all three sequencing modalities are even more limited (14). Here, we address common considerations in choosing an NGS method (Table 2). Additionally, we review studies within the past five years that have directly compared NGS methods in humans (Table 3).

16S rRNA, shotgun metagenomic, and RNA sequencing can all be used to determine what bacteria are present in a microbiome; however, the latter two also detect members of other domains such as fungi and parasites, as well as viruses. Only RNA sequencing examines RNA viruses. With respect to taxonomic resolution, an overarching finding of the studies that have compared these methods is that phylum designations are comparable (39); however, 16S rRNA sequencing tends to offer less resolution and sensitivity for detecting changes at the species level and cannot detect strain-level changes (13, 34, 43). For example, Jovel and colleagues conducted parallel 16S rRNA and shotgun metagenomic sequencing on mock bacterial populations with defined consortia and found that the 16S rRNA method and software pipelines (Quantitative Insights Into Microbial Ecology [QIIME], refs. 44, 45; and mothur, ref. 46) effectively resolved sequences to the genus level, but shotgun metagenomic sequencing resulted in improved genus- and species-level classification (36). This finding has been replicated in other human studies (Table 3 and refs. 14, 15, 35–40). Interestingly, Drewes et al. compared 16S rRNA analysis pipelines and found that the Resphera Insight high-resolution taxonomic assignment tool (Resphera Biosciences; refs. 47–49) better characterized species-level differences using human colon cancer samples compared with other 16S rRNA sequencing pipelines (50). To our knowledge, no studies, as yet, have directly compared the Resphera Insight (47–49) pipeline species classification with that of shotgun metagenomic sequencing. Very recently, the Kraken pipeline for shotgun metagenomic analysis was expanded to enable 16S rRNA analysis, and results show that it is more accurate and up to 300 times faster than QIIME (51). However, QIIME includes a wealth of other helpful tools, making it more a stand-alone "complete" package. For users sophisticated enough to mix-and-match packages, Kraken could replace the core QIIME step of 16S read assignment.

A functional profile cannot be directly obtained from 16S rRNA sequencing, because the method only characterizes sequences from one essential gene. Methods like PICRUSt (Phylogenetic Investigations of Communities by Reconstruction of Unobserved States; ref. 52) or PICRUSt2 (53) and Tax4Fun (54) or Tax4Fun2 (55) aim to predict functional profiles of bacteria based on 16S rRNA data. However, the success of these methods, when compared with functional potentials obtained via shotgun metagenomics, varies with the 16S gene primers used for amplification (35, 36). Conversely, shotgun metagenomics and RNA sequencing consider all the microbial DNA and RNA; thus it is possible to more comprehensively predict the functional potential. Importantly, a distinct difference between shotgun metagenomics and RNA sequencing is that shotgun metagenomics provides a random selection of all genes encoded by the microbes (predictive functional potential) whereas RNA sequencing identifies which genes are actively being transcribed (active functional profile).

Other considerations in pursuing an NGS method and analyses include host contamination, false positives, bias, and post-sequencing computational requirements. There is less risk of host contamination in 16S rRNA sequencing compared with other NGS methods because the gene being amplified and

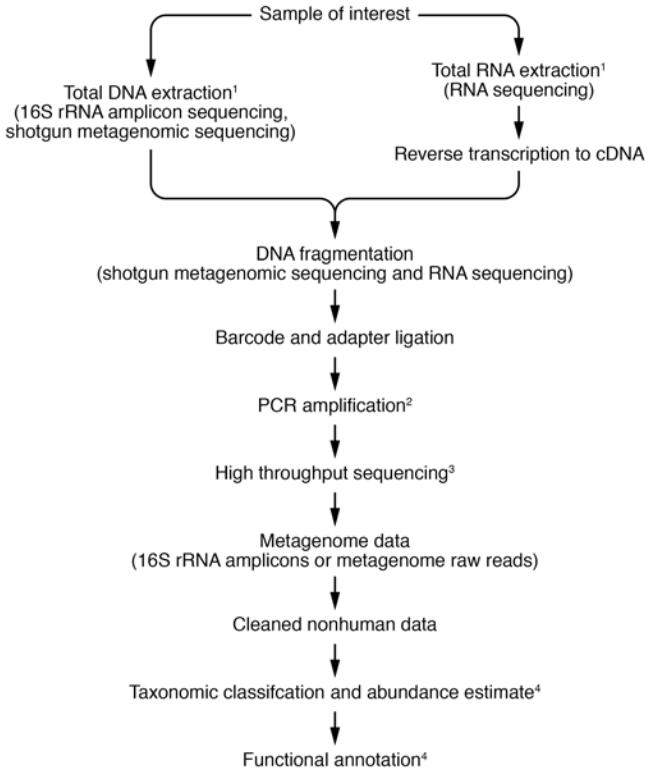

Figure 2. NGS implementation. Overview of key steps in 16S rRNA gene sequencing, shotgun metagenomic sequencing, and RNA sequencing processes. [1]Host DNA or RNA depletion can be performed (optional steps). [2]PCR amplification is used to amplify bacterial 16S rRNA gene variable regions (16S rRNA amplicon sequencing) or random cDNA fragments resulting from RNA reverse transcription for RNA sequencing. DNA-based shotgun metagenomic sequencing is optimally done without use of PCR amplification to avoid introduction of PCR-associated experimental bias. However, in samples with low DNA quantities, PCR amplification of the DNA library is sometimes used. [3]Commonly Illumina-based sequencing chemistry (33). [4]The taxonomic and functional analyses of NGS data are complex and make use, most often, of software available in the public domain.

sequenced (i.e., the 16S rRNA gene) is specific to bacteria. With 16S rRNA sequencing, there is also a lower risk of false positives due to extensive reference databases and computational error correction tools; however, the risk of false positives increases with decreasing sample biomass (33). Conversely, there is a higher risk of bias with 16S rRNA sequencing because of primer-dependent PCR amplification bias and differences between the variable regions, as discussed above (17–19). Importantly, one must also consider the computational expertise and analysis required after sequencing. Currently, 16S rRNA sequencing bioinformatics analysis is less of an undertaking than either shotgun metagenomics or RNA sequencing, as there are fewer data (i.e., sequencing output from one gene versus all genes) as well as several publicly available and user-friendly platforms, like QIIME (44, 45) and mothur (46). This makes 16S rRNA sequencing more accessible to researchers with beginner- and intermediate-level bioinformatics experience (33). For projects directed at detection of specific taxa, pilot data using mock microbial communities can guide experimental choices (e.g., primer sets and/or estimation of read numbers or sequencing depth [see below] needed for taxon identification).

Finally, cost must be considered for any project and is arguably one of the most important factors in what type of NGS to initially perform. The differences in cost between the methods relate to the amount and depth of sequencing. Sequencing depth refers to the number of times a certain nucleotide base is represented in the sequencing reads for a given sample (56). Typically, shotgun metagenomics and RNA sequencing analyses require much more sequence data than 16S rRNA sequencing, resulting in their higher costs. However, a recent study by Laudadio and colleagues suggests that shotgun metagenomics, at lower sequencing depths, is comparable in price to 16S rRNA sequencing and still identifies more species (38). Notably, this study did not consider other inherent NGS costs, including computational burden and data storage.

In summary, the use of the 16S rRNA gene as a phylogenetic marker is efficient and cost effective (52); however, it is subject to biases that other microbiome characterization methods are not (i.e., choice of hypervariable regions and primer-dependent PCR amplification) and can thus result in significant variance in the determined microbial composition of a sample. Additionally, 16S rRNA sequencing is commonly limited to taxonomic classification at the genus level or above (36), as horizontal transfer of the 16S rRNA locus and the existence of multiple bacterial species and strains that are more than 97% similar can prevent more nuanced classification (35, 43). Finally, 16S rRNA analysis provides limited predicted functional information (14, 52). Conversely, shotgun metagenomics and RNA sequencing are more expensive than 16S rRNA sequencing but offer far broader taxonomic coverage (i.e., species- and strain-level resolution), more accurate functional profiling, and the possibility of detecting previously unknown species and strains of microbes (36). Although shotgun metagenomic and/or RNA sequencing undoubtedly provides more information, determining which approach is appropriate depends on the question(s) being asked. For instance, if you want to identify the dominant bacteria in a sample, 16S rRNA sequencing is likely the better method owing to the lower cost and bioinformatics burden (42). We present comparisons herein not to suggest that one sequencing method or protocol is best for all projects but rather to assist readers in selecting the best protocol for their projects.

Technical and individual laboratory issues: sources of variability. There are multiple parameters to consider regarding sample collection and processing, because variabilities in any of these steps can alter NGS data. First, the investigator must choose the type of sample for NGS sequencing. Although fecal samples and body fluids are easier to collect and permit serial sampling, intraluminal fecal samples or tissue samples may provide representative regional colon or site-specific microbiome characterization. Storage conditions can further impact NGS results, and thus this information should be reported. The gold standard is immediate freezing of samples and storage at –80°C (57); however, samples can also be preserved chemically using solutions such as DNA/RNA Shield (Zymo Research) (58).

The first step in sample processing is DNA or RNA extraction, and this step is responsible for the majority, but not all, of experimental variability in microbiome analysis according to the MicroBiome Quality Control project (59). Numerous commercially

Table 2. Comparisons of common microbiome sequencing methods

	Method		
	Amplicon (16S, 18S, ITS[A])	Shotgun metagenomics	RNA sequencing
What is sequenced?	DNA coding for the 16S, 18S ribosomal subunit or ITS	Host and microbial DNA	Host and microbial RNA
What is the taxonomic resolution?	Phylum–genus, sometimes species	Species–strains	Species–strains
What is the taxonomic coverage?	Bacteria, archaea (16S); eukaryotes (18S)	Bacteria, archaea, eukaryotes, DNA viruses	Bacteria, archaea, eukaryotes, DNA and RNA viruses
Are appropriate reference databases available?	Over 3 million 16S gene sequences from humans and environmental sources are available	Over 100,000 genomes with a bias toward human microbiomes	Over 100,000 genomes with a bias toward human microbiomes
Does host contamination occur?	Limited	Yes, but can be mitigated by host DNA/rRNA depletion methods	Yes, but can be mitigated by host DNA/rRNA depletion methods
Can sequencing data yield a functional profile?	Not directly, but the functional profile can be predicted computationally	Yes, with appropriate computational expertise	Yes, with appropriate computational expertise
What is the minimum input for detection?	10 copies	1 ng[B]	1 ng[B]
What is the potential for false positives?	Lower due to extensive reference databases and error correction tools	Higher due to host DNA contamination of draft genomes	Higher due to host RNA/DNA contamination of draft genomes
What is the potential for bias?	Medium to high due to a dependence on primers, a targeted variable region, and PCR amplification	Lower due to the untargeted nature of the methodology	Lower due to the untargeted nature of the methodology
What level of computational skills is required?	Beginner–intermediate	Intermediate–advanced	Intermediate–advanced

[A]16S rRNA amplicons identify bacteria; 18S rRNA amplicons and internal transcribed spacer (ITS) sequences are most often used to identify fungi or parasites. [B]Although 1 ng is considered the minimum input, many sequencing facilities require at least 20 ng. Adapted from Zymo Research (140) and *Protein and Cell* (33).

available kits exist for DNA extraction, including from Covaris, Qiagen, Zymo Research, and others. Typically samples are homogenized, but protocols vary substantially from laboratory to laboratory (59). Although there is not yet a globally accepted gold-standard protocol for DNA or RNA extraction, it is critical that all samples be processed in the same manner. Furthermore, it is strongly recommended that negative controls be processed to better assess the comparability of different NGS runs, normalize across separate NGS runs to limit batch effects, identify kit-specific contaminants, and determine whether the detection of low-abundance microbes in a sample are of biologic interest or, more likely, represent contaminants. Examples of controls include (a) storage buffer (e.g., DNA/RNA Shield); (b) DNA extraction kit components; and (c) a community standard containing known species at known quantities (e.g., Zymo Microbial Community Standard [D6300] and Zymo Microbial Community Standard II Log Distribution [D6310]).

For 16S rRNA sequencing, the PCR amplification step is also a source of variability. As discussed, there are nine hypervariable regions in the 16S rRNA gene, and available primer sets typically amplify only a subset of these regions. Thus, the performance characteristics of the primer set chosen will influence the number of the analyzable reads (60) as well as the results of the analysis (61). For example, one study reports that the V4 primer set yields significantly more *Bacteroides* and lower Firmicutes reads than other primer sets tested; this is particularly notable given that the Bacteroidetes/Firmicutes ratio is a commonly reported metric (60).

The results of sequencing itself also vary with different equipment, and thus, ideally, all samples are sequenced using the same sequencing platform (e.g., Illumina MiSeq, NovaSeq). Finally, a wide variety of bioinformatics pipelines are available, for both 16S rRNA and shotgun metagenomics data, and the choice of computational and statistical methods can have a critical effect on outcomes and conclusions (36, 59, 62), including the risk of reporting false associations and of missing true ones. While a full review of computational methods and their relative strengths and weaknesses is beyond the scope of our discussion, Liu et al. (33) provided a recent review covering dozens of methods.

Overall, variability in any of the steps of NGS sequencing (e.g., sample type, sample storage, DNA extraction, PCR amplification, sequencing technology, read length, and/or bioinformatics analysis) can lead to data variability. There is generally not a "right" answer as to the best method or approach. The most important principle is that all samples be treated the same to facilitate meaningful comparisons between samples in the same study. However, as discussed in the next section, great care must be taken in comparing results between different studies, as these variables may differ.

Challenges of rigor, reproducibility, and reporting. Microbiome science is complex, cutting across many scientific fields, including microbiology, epidemiology, biology, computational science, genomics, and biostatistics. This complexity and the rapid evolution of approaches within the field have led to the reporting of disparate findings between studies investigating seemingly similar patient populations. Thus, increasing attention is now directed to developing well-curated and validated databases that are critical for accurate analyses, and providing guidance for the consistent conduct and reporting of study design, methods, and results of

Table 3. Recent comparisons of NGS methods for microbial taxonomic classification and functional profiling in human samples

Ref.	Sample type and population studied (*n*)	Methods	Relevant findings
35	Human fecal sample (*n* = 1)	16S rRNA vs. shotgun metagenomic sequencing • Equipment[A]: Illumina MiSeq and HiSeq 2000 • V1–V3 16S rRNA gene primers	• 16S rRNA sequencing detected notable phylum-level changes in Bacteroidetes and Actinobacteria compared with shotgun metagenomics. Firmicutes and Proteobacteria abundances were similar between methods. • Shotgun metagenomics detected approximately twice as many species as 16S rRNA sequencing (Pearson's correlation coefficient 0.6). • The reference database impacted species-level calls for both methods. • α-Diversity[B] was lower for 16S rRNA sequencing than for shotgun metagenomics.
15	Geriatric human fecal samples (*n* = 6)	16S rRNA vs. shotgun metagenomic sequencing • Equipment[A]: Illumina MiSeq, HiSeq, and Ion PGM • V1–V2 and V4–V5 16S rRNA gene primers	• Illumina MiSeq and HiSeq resulted in more reads and detected more species than Ion PGM. Illumina HiSeq shotgun metagenomics identified the most species. • Samples clustered by primer type or sequencing platform as opposed to sample donor. This observation was more pronounced for 16S rRNA data. For shotgun metagenomics, MetaPhlAn (141), followed by Kraken (142), produced the least biased results.[C]
36	Mock bacterial populations (*n* = 3); kefir (*n* = 1); mouse (*IL-10*$^{-/-}$) fecal samples (*n* = 3); human fecal and ileal tissue samples from patients with *C. difficile* (*n* = 1) or Crohn's disease (*n* = 2) and healthy controls (*n* = 3)	16S rRNA vs. shotgun metagenomic sequencing • Equipment[A]: Illumina MiSeq • V4 16S rRNA gene primers	• In general, regardless of the type of sample analyzed, shotgun metagenomic sequencing (using BLAST [ref. 143], MEGAN [ref. 144], MetaPhlAn [ref. 141] for analysis) identified more taxa and, in particular, species than 16S rRNA sequencing (using QIIME [refs. 44, 45] and mothur [ref. 46] for analysis). MetaPhlAn was the fastest and most precise. • Increased sampling depth and decreased sample complexity improved method concordance and taxon identification. However, increased sampling depth may also enhance detection of low-level environmental contaminants. • α-Diversity[B] was similar between methods. • Predicted functional profile concordance varied between PICRUSt (52) for 16S rRNA data and MEGAN5 (144) for shotgun data.
14	Human fecal sample (*n* = 1) with bacterial spike-ins	16S rRNA vs. shotgun metagenomic vs. meta-total RNA sequencing[D] • Equipment[A]: Illumina HiSeq 2500 • V4–V5 16S rRNA gene primers	• Meta-total RNA sequencing[D] resulted in higher α-diversity[B] than 16S rRNA sequencing and shotgun metagenomics. Shotgun metagenomics resulted in higher α-diversity than 16S rRNA sequencing. • Meta-total RNA sequencing detected more genera than 16S rRNA and shotgun metagenomic sequencing. 16S rRNA sequencing detected more genera than shotgun metagenomics.
38	Pediatric human fecal samples: patients with Crohn's disease (*n* = 3), healthy controls (*n* = 3)	16S rRNA vs. shotgun metagenomic sequencing • Equipment[A]: Illumina MiSeq and HiSeq 2500 • V3–V4 16S rRNA gene primers	• Overall, shotgun metagenomics (using MetaPhlAn [ref. 141] for analysis) identified 3 times as many species as 16S rRNA sequencing (using MICCA [ref. 145] and QIIME 2 [refs. 44, 45] for analysis).[E] • Shotgun metagenomics identified more species than 16S rRNA sequencing at every sequencing depth tested. The lowest depth tested was approximately 1 million paired-end reads.
40	Human infant fecal samples (*n* = 338)	16S rRNA vs. shotgun metagenomic sequencing • Equipment[A]: Illumina MiSeq and NextSeq 550 • V4–V5 16S rRNA gene primers	• Shotgun metagenomics identified more species, but fewer taxa at the family and genus levels, than 16S rRNA sequencing. Consistent with this observation, 16S rRNA sequencing yielded greater α-diversity[B] at the genus level than shotgun metagenomics. • α-Diversity increased with sequencing depth. • At the genus level, β-diversity[F] was concordant between methods.

[A]Equipment used for 16S rRNA, shotgun metagenomic sequencing, or RNA sequencing as indicated. [B]α-Diversity (within-sample diversity) can be measured using multiple indices that reflect sample richness (number of taxa) and/or evenness (abundance). Commonly used indices in the cited papers are the Shannon diversity index (146) and Simpson index (147). These account for richness and evenness of taxa. [C]The term "least biased results" indicates that samples clustered more by sample than method. [D]"Meta-total RNA sequencing" refers to a method described by Cottier and colleagues (14) that is akin to shotgun sequencing of total RNA but may require lower sequencing depth than shotgun metagenomics. [E]It is important to note that the default analysis parameters used by Laudadio and colleagues (38) for QIIME2 (44, 45) and MICCA (145) assigned a taxon if a single confidently assigned read was identified. In contrast, taxon assignment for shotgun metagenomics analysis was more conservative. In individual papers, these parameters can be modified in the analyses and may impact the results. [F]β-Diversity (between-sample diversity) is commonly measured by principal coordinate analysis using the Bray-Curtis distance.

microbiome research. In 2018, Schloss provided a thoughtful and pragmatic essay for translational researchers to consider the threats to rigor, reproducibility, and generalizability within microbiome research (63). Others have called for a centralized robust curated data repository for microbiome data adherent to FAIR (findable, accessible, interoperable, and reproducible) principles (64). Consistent with this need, the FDA has established an evolving quality-controlled and highly curated public microbial reference database (FDA-ARGOS) for microbiome research (65), although this database is still relatively small. Most recently, the STORMS (Strengthening the Organization and Reporting of Microbiome Studies) 17-point Microbiome Reporting Checklist was proposed as a guide for researchers, reviewers, and readers for the presentation, assessment, and understanding of microbiome research across studies (66, 67). Although STORMS was developed through a strong iterative process, it is based on the analysis of only one paper and has been minimally used to date (66). Nonetheless, previous reporting guidelines — e.g., CONSORT (Consolidated Standards of Reporting Trials) — improved the quality of clinical trial reporting (66), and such results support calls for more structured microbiome research reporting. Improvement of microbiome science communication and of the ability to cross-compare studies is essential for human microbiome studies to yield progress in applying microbiome science to patient care.

Essential considerations in collection of clinical metadata. Beyond the complexities of designing the laboratory, computational, and statistical approaches to NGS-driven human studies, the investigator must also consider what and how much clinical metadata to collect. Age, sex, and geography are fundamental as each impacts microbiome composition and likely function (68, 69). However, given the interindividual variability in the microbiome (70) and disease-associated data (discussed below), more nuanced considerations of individual exposures, both current and over time, may be needed. These include environmental exposures associated with migration (71), diet and food additives (72), and antibiotic and non-antibiotic medications (73, 74). While genetic impacts on the human microbiome have been downplayed in recent literature (75), this is likely short-sighted, as we do not yet understand how microbial communities function, and data suggest that select members of the microbiome serve as functional drivers that intersect with host genes to regulate clinical outcomes (76–78). This broad field of human exposures that impact health and disease is termed the "exposome" and, while impossible to fully capture in most studies, deserves careful thought in study design, data accrual, and interpretation (79).

Relevance of NGS to clinical and translational research

Although extremely useful in investigating disease mechanisms that may inform human translational research, herein, we will not consider the enticing but likely overinterpreted rodent microbiome studies (80). Instead, given the breadth of available data, we focus on a few illustrative examples of human microbiome analyses to indicate the robust impact that NGS-derived microbiome data can have on our thinking about human diseases; such results implicate the potential for human microbiome science to impact clinical care (81).

Nutrition and metabolism. In the very active area of investigation concerning nutrition and metabolism (82), we provide a few seminal observations that may help guide considerations in NGS microbiome and implementation research. Only key highlights from each paper are presented, and the reader is referred to the individual publications for further details. In 2011, Wu and colleagues provided human data strongly linking diet and gut microbiome composition (83). This feeding study identified that a diet change (low fat/high fiber versus high fat/low fiber) led to detectable gut microbiome changes within 24 hours, without perturbation of overall compositional microbiome structure. These data provide insight into rapid diet-dependent microbiome shifts, but suggest that long-term diet is key to overall gut microbiome structure and likely function. O'Keefe and colleagues demonstrated that a mere 2-week switch in diet, from a US-based Western diet (high fat/low fiber) to a rural African diet (low fat/high fiber) or vice versa, led to remarkable reciprocal changes in mucosal inflammation and proliferation as well as metabolic health indicators in African American and rural South African populations (84). In a very detailed study of individual diet and health impact, Zeevi and colleagues identified, unexpectedly, the wide variability in human diet metabolic processing and physiologic impact. Namely, postprandial glycemic response to identical meals and combination foods like pizza varies dramatically between individuals, but can be predicted using a machine learning algorithm that integrates microbiome NGS data and other inputs (85). For example, pizza may not significantly alter the metabolism of one person but may induce hyperglycemia in another. Lastly, the Gordon laboratory, based on at least a decade of investigations integrating rodent-based experimental and human studies, provided the first microbiota-directed complementary food prototype, termed MDCF-2. This proof-of-principle, prospective, randomized study was conducted in a population of children with moderate acute malnutrition in Bangladesh. It yielded strong evidence that a diet composed of locally sourced foods could improve weight-for-length *z* scores in 3 months but also identified that the *z* scores declined quickly after MDCF-2 supplementation withdrawal (86). It remains to be determined whether this diet, locally sourced from Bangladeshi foods, will yield similar results in other global locales and/or whether the MDCF development process can be streamlined to yield products able to promote global nutritional health equity.

The skin in health and disease. The skin microbiota is essential for protection against invading pathogens. One remarkable aspect of the skin microbiome is its regional diversity whereby the local (e.g., ear versus navel) microbiome varies, possibly because of differential environmental exposures (70). The accessibility of the skin and ease of sampling have fostered exemplary longitudinal studies (needed in other areas of microbiome science) of conditions like atopic dermatitis. This work provides insight into the skin microbiome and disease-associated bacterial strain fluctuations, although disease mechanisms require further study (87, 88).

Early life exposures. Microbiome analyses of birth cohorts and early-life exposures underpin our fundamental understanding of development (89), exposure impacts (e.g., cesarian versus vaginal delivery [ref. 90], antibiotics [ref. 91]), and disease onset (e.g., childhood asthma, atopy [ref. 92]). This work highlights that the

early-life fecal compositional assembly and metabolome associate with the emergence of childhood atopy and asthma years later, in part because of immune development dysregulation (92, 93). It also underpins the importance of microbiome analyses for the prediction of disease development in additional, large, longitudinal birth cohort studies. One example is a study of 100,000 mother-baby pairs in the Greater Bay Area in China, led by the Faculty of Medicine at the Chinese University of Hong Kong (64).

Defining outbreak transmission, source, and pathogenicity. NGS studies, both whole-genome and metagenomic sequencing, along with detailed epidemiologic analyses have been instrumental in tracking and identifying the source of multi-drug-resistant pathogens, persistent even over extended time periods in a hospital (94, 95). Pathogen identification enabled interventions to eliminate the infection source and understand hospital spread. Furthermore, NGS studies of outbreak human *Burkholderia* strains, isolated from individuals with cystic fibrosis, led to the identification of bacterial genes promoting this bacterium's human host adaptation and virulence (96). Identifications such as these offer insights for new therapeutic targets.

Impact of NGS on disease diagnosis. A clinical benefit of microbiome NGS may be to predict disease risk, akin to the established use of human genome NGS to identify disease risk. Microbiome NGS to predict disease risk is not yet validated for any disease, but progress is occurring. For example, as described above, longitudinal studies in children have begun to link microbes to risk for onset of asthma and atopic conditions (92, 93). Another example is the use of the colon microbiome (i.e., colon mucosal or fecal samples) to predict colorectal cancer (CRC) risk. To date, although metagenomic analyses detect microbial communities reflective of CRC, detection of communities reflective of precancerous lesions (e.g., colonic polyps) is limited (97, 98). Similarly, blood-based transcriptomes best detect advanced-stage cancers (99). This suggests that NGS methods need further development to detect early-stage disease when intervention may enhance patient prognosis (100).

The clinical microbiology laboratory is beginning to use microbial NGS methods for disease diagnosis, particularly to identify potential infectious etiologies of chronic illnesses. Use of NGS has emerged to define undiagnosed CNS infections (12, 101, 102), respiratory pathogens (103), and other difficult-to-diagnose or undiagnosed infectious diseases. Metagenomic sequencing is attractive for detecting suspected, but undiagnosed, infections because nucleic acid analyses can, theoretically, detect bacteria, viruses, fungi, and parasites. This extensive detection potential could limit the numerous tests required to assess a broad array of putative pathogens in patients without diagnoses. Hurdles include differentiating colonization from infection, limiting contaminants, developing efficient, clinical sample–specific methods, achieving analytical standardization, and continually working to improve data security to protect patient privacy. Cost is another hurdle, as NGS method validation can be very expensive (12).

Research is needed to define how use of microbiome NGS can advance patient care, solve the source of outbreaks (e.g., *Klebsiella* [ref. 94] and *Sphingomonas* [ref. 95]), and identify and characterize emerging pathogens (e.g., SARS-CoV-2 [ref. 104]). The most difficult challenge for clinicians and translational scientists is the interpretation of NGS data for clinical application (12, 105, 106).

Development of therapeutics from NGS and the microbiome

To date, there are no FDA-approved therapeutics based on NGS or derived from the human microbiota or microbiome. Nonetheless, this is a rich area of research, and we highlight some important and ongoing work in the next sections.

Whole community transfer: fecal microbiota transplantation. Fecal microbiota transplantation (FMT) is an ancient therapy (107), having been employed as early as the 4th century BCE in China, and has been proposed as the method most likely to succeed in manipulating pathophysiologically complex diseases (108). FMT is used primarily for the treatment of *Clostridioides difficile* disease but is also being explored, with variable clinical outcomes, as a therapeutic for other gastrointestinal (109, 110) and non-gastrointestinal diseases (111, 112). Key limitations of FMT, as currently used, are its inherent lack of quality control and imprecision, combined with our weak understanding of the microbes and mechanisms by which FMT may confer benefit. Furthermore, enthusiasm for the use of FMT is now more restrained with the emergence of SARS-CoV-2 (live virus is present in feces; refs. 113, 114), safety concerns (115, 116), deaths (117), FDA warnings (118), and more stringent screening requirements (119). Recent analysis, with improved delineation of variables impacting FMT, indicates that its outcomes, even with *C. difficile* disease, may not be as robust as previously suggested by the case report literature (116, 120). However, promising microbiome community–based quality-controlled FMT products (e.g., SER-109, in ECOSPOR clinical trials; and RBX2660, in PUNCH clinical trials; ref. 121), based, at least in part, on NGS, are being studied in prospective, randomized clinical trials (SER-109, ref. 122; RBX2660, refs. 123, 124; >400 studies at ClinicalTrials.gov, accessed November 27, 2021). Most recently, a phase III, double-blind, placebo-controlled trial of SER-109, an oral microbiome therapeutic composed of human stool–derived live Firmicutes bacterial spores, reported efficacy superior to that of placebo in lowering rates of *C. difficile* recurrence in all age groups studied (recurrence, SER-109 vs. placebo, 12% vs. 40%; relative risk, 0.32; 95% confidence interval, 0.18–0.58, $P < 0.001$). The safety profile was similar to placebo (125). FMT and microbial replacement products require more study to understand the mechanisms, microbes, and durability by which these therapeutics alter gut microbiome function and drive clinical outcomes (126).

Additions to the host microbiome: prebiotics and probiotics. Both untargeted and targeted approaches to modulate the function of the microbiome are being studied, each of which utilizes and/or requires NGS to assess impact. The first and most prevalent untargeted example over time has been the ingestion of over-the-counter prebiotics and probiotics. Prebiotics are substrates (e.g., fiber) that are consumed by gut microbes, whereas probiotics are live organisms ingested to confer health benefits. Because most prebiotics and probiotics are not subject to regulatory oversight, the products can be highly variable (127). Recent data, for example, have demonstrated a lack of benefit of the most commonly used probiotic globally, *Lactobacillus rhamnosus* (LGG or R0011), studied with or without *L. helveticus* R0052, in childhood diarrhea (128, 129). Furthermore, the colonization by and effect of probiotic strains appear to vary significantly between individuals (82). Nonetheless, another recent study, using fecal microbiome analyses of

individuals in rural Thailand, identified that *Bacillus*, a spore-forming bacterium, was associated with reduced human *Staphylococcus aureus* colonization, a cause of systemic antibiotic-resistant infections. This outcome was ascribed to *Bacillus* production of lipopeptides (fengycins) that inhibit *S. aureus* quorum-sensing mechanisms (130). Thus, development of rational precision probiotics based on NGS and microbiome research is likely feasible and is another key microbiome research opportunity.

Preclinical targeted NGS-linked tactics to modulate the microbiome. Multiple approaches, under development, may allow for precision manipulation of the gut microbiome and have potential to impact local and/or systemic disease processes. These include (a) inhibition of gut bacterial enzymes to modify metabolic capabilities; (b) selective bacteriophage-mediated depletion of disease-inducing or undesirable bacterial strains; (c) gut colonization with engineered strains that deliver a therapeutic payload; and (d) direct genetic modification of the in situ microbiome (131). Excitingly, human proof-of-principle studies already exist for some of these approaches. For example, bacteriophages have been successfully used to treat systemic antimicrobial-resistant infections (132, 133), and an engineered, oral *E. coli* Nissle strain promoted arginine synthesis from ammonia in healthy volunteers (134) and is being developed as a potential treatment for hyperammonia conditions (e.g., hepatic encephalopathy).

The microbiome as a source of new drugs. Many antibiotics, including penicillin, are natural products or their derivatives. As an extension of this prior success, high-dimensional, multicomponent screening strategies have recently been used to identify antimicrobials with novel mechanisms of action from uncultured soil or marine microbiome members (135, 136). Whether these microbiome derivatives will be successful in humans remains to be tested.

Conclusions

Microbiome science is moving toward microbiome precision medicine, but we still lack sufficient clinical data, either cross-sectional or longitudinal, to apply the science to human health or disease with confidence. Investigative needs include integrative human-centric microbiome studies, broader and more consistent integration of the exposome, a better understanding of the putative unique contributions of different analytical approaches, and cross-validation of data sets between studies of disease processes and populations. Ultimately, NGS results will be complemented with non-NGS methods such as metabolomics and proteomics to understand microbial functions in health and disease. Critically, data must be interpreted with consideration of clinical plausibility. Furthermore, and in parallel, investigators should be explicit about gaps in knowledge or new directions for additional clinical studies.

Application of NGS data and microbiome investigations in clinical medicine is in its infancy, and thus contains both promise and uncertainty. The current paucity of carefully designed prospective and longitudinal human studies highlights a rich opportunity for clinical translational scientists. By leveraging the cross-disciplinary nature and complexities of microbiome science, we can advance our understanding of disease development, progression, diagnosis, and therapy, ultimately benefiting the health of patients.

Acknowledgments

The authors thank their current and former laboratory members for shaping the insights shared. This work was funded by NIH grants R01-CA196845 (to CLS), R35-GM130151 (to SLS), R01-HG006677 (to SLS), and R56DK107726 (to JLP); the Bloomberg-Kimmel Institute for Cancer Immunotherapy (to CLS); a Cancer Grand Challenges OPTIMISTICC team grant (A27140) funded by Cancer Research UK (to CLS); and the Johns Hopkins School of Medicine (to CLS).

Address correspondence to: Cynthia L. Sears, Johns Hopkins University School of Medicine, 1550 Orleans Street, CRB2 Building, Suite 1M.05, Baltimore, Maryland 21231, USA. Phone: 410.614.8378; Email: csears@jhmi.edu.

1. Sender R, et al. Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol*. 2016;14(8):e1002533.
2. Smith I. Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. *Clin Microbiol Rev*. 2003;16(3):463–496.
3. LaFond RE, Lukehart SA. Biological basis for syphilis. *Clin Microbiol Rev*. 2006;19(1):29–49.
4. Gevers D, et al. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe*. 2014;15(3):382–392.
5. Zhou M, et al. Investigation of the effect of type 2 diabetes mellitus on subgingival plaque microbiota by high-throughput 16S rDNA pyrosequencing. *PLoS One*. 2013;8(4):e61516.
6. Davis CD. The gut microbiome and its role in obesity. *Nutr Today*. 2016;51(4):167–174.
7. Sears CL, Garrett WS. Microbes, microbiota, and colon cancer. *Cell Host Microbe*. 2014;15(3):317–328.
8. Finlay BB. Are noncommunicable diseases communicable? *Science*. 2020;367(6475):250–251.
9. Neville BA, et al. Commensal Koch's postulates: establishing causation in human microbiota research. *Curr Opin Microbiol*. 2018;42:47–52.
10. Falkow S. Molecular Koch's postulates applied to bacterial pathogenicity—a personal recollection 15 years later. *Nat Rev Microbiol*. 2004;2(1):67–72.
11. Rappé MS, Giovannoni SJ. The uncultured microbial majority. *Annu Rev Microbiol*. 2003;57(1):369–394.
12. Simner PJ, et al. Understanding the promises and hurdles of metagenomic next-generation sequencing as a diagnostic tool for infectious diseases. *Clin Infect Dis*. 2018;66(5):778–788.
13. Durazzi F, et al. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Sci Rep*. 2021;11(1):3030.
14. Cottier F, et al. Advantages of meta-total RNA sequencing (MeTRS) over shotgun metagenomics and amplicon-based sequencing in the profiling of complex microbial communities. *NPJ Biofilms Microbiomes*. 2018;4(1):1–7.
15. Clooney AG, et al. Comparing apples and oranges?: next generation sequencing and its impact on microbiome analysis. *PLoS One*. 2016;11(2):e0148028.
16. Lane DJ, et al. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A*. 1985;82(20):6955–6959.
17. Chakravorty S, et al. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods*. 2007;69(2):330–339.
18. Rintala A, et al. Gut microbiota analysis results are highly dependent on the 16S rRNA gene target region, whereas the impact of DNA extraction is minor. *J Biomol Tech*. 2017;28(1):19–30.
19. Sirichoat A, et al. Comparison of different hypervariable regions of 16S rRNA for taxonomic profiling of vaginal microbiota using next-generation sequencing. *Arch Microbiol*. 2021;203(3):1159–1166.
20. Johnson JS, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun*. 2019;10(1):5029.
21. Xia Y, et al. *Statistical Analysis of Microbiome Data with R*. Springer; 2018.
22. Cole JR, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res*. 2009;37(suppl 1):D141–D145.

23. Quast C, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41(d1):D590–D596.
24. DeSantis TZ, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol*. 2006;72(7):5069–5072.
25. Tyler AD, et al. Analyzing the human microbiome: a "how to" guide for physicians. *Am J Gastroenterol*. 2014;109(7):983–993.
26. Schloss PD. Amplicon sequence variants artificially split bacterial genomes into separate clusters. *mSphere*. 2021;6(4):e0019121.
27. Callahan BJ, et al. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J*. 2017;11(12):2639–2643.
28. Staley JT. The bacterial species dilemma and the genomic-phylogenetic species concept. *Philos Trans R Soc Lond B Biol Sci*. 2006;361(1475):1899–1909.
29. O'Leary NA, et al. Reference Sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44(d1):D733–D745.
30. Benson DA, et al. GenBank. *Nucleic Acids Res*. 2018;46(d1):D41–D47.
31. Wattam AR, et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res*. 2014;42(d1):D581–D591.
32. Aurrecoechea C, et al. EuPathDB: the eukaryotic pathogen genomics database resource. *Nucleic Acids Res*. 2017;45(d1):D581–D591.
33. Liu Y-X, et al. A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein Cell*. 2021;12(5):315–330.
34. Shah N, et al. Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. *Pac Symp Biocomput*. 2011;165–176.
35. Ranjan R, et al. Analysis of the microbiome: advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem Biophys Res Commun*. 2016;469(4):967–977.
36. Jovel J, et al. Characterization of the gut microbiome using 16s or shotgun metagenomics. *Front Microbiol*. 2016;7:459.
37. Eloe-Fadrosh EA, et al. Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. *Nat Microbiol*. 2016;1(4):1–4.
38. Laudadio I, et al. Quantitative assessment of shotgun metagenomics and 16S rDNA amplicon sequencing in the study of human gut microbiome. *OMICS*. 2018;22(4):248–254.
39. Rausch P, et al. Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome*. 2019;7(1):133.
40. Peterson D, et al. Comparative analysis of 16S rRNA gene and metagenome sequencing in pediatric gut microbiomes. *Front Microbiol*. 2021;12:1651.
41. Regalado J, et al. Combining whole-genome shotgun sequencing and rRNA gene amplicon analyses to improve detection of microbe-microbe interaction networks in plant leaves. *ISME J*. 2020;14(8):2116–2130.
42. Brumfield KD, et al. Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. *PLoS One*. 2020;15(2):e0228899.
43. Poretsky R, et al. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One*. 2014;9(4):e93827.
44. Caporaso JG, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7(5):335–336.
45. Bolyen E, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37(8):852–857.
46. Schloss PD, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75(23):7537–7541.
47. Daquigan N, et al. Early recovery of *Salmonella* from food using a 6-hour non-selective pre-enrichment and reformulation of tetrathionate broth. *Front Microbiol*. 2016;7:2103.
48. Ottesen A, et al. Enrichment dynamics of Listeria monocytogenes and the associated microbiome from naturally contaminated ice cream linked to a listeriosis outbreak. *BMC Microbiol*. 2016;16(1):275.
49. Abernethy MG, et al. Urinary microbiome and cytokine levels in women with interstitial cystitis. *Obstet Gynecol*. 2017;129(3):500–506.
50. Drewes JL, et al. High-resolution bacterial 16S rRNA gene profile meta-analysis and biofilm status reveal common colorectal cancer consortia. *NPJ Biofilms Microbiomes*. 2017;3(1):1–12.
51. Lu J, Salzberg SL. Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. *Microbiome*. 2020;8(1):124.
52. Langille MGI, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol*. 2013;31(9):814–821.
53. Douglas GM, et al. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol*. 2020;38(6):685–688.
54. Aßhauer KP, et al. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics*. 2015;31(17):2882–2884.
55. Wemheuer F, et al. Tax4Fun2: prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene sequences. *Environ Microbiome*. 2020;15(1):11.
56. Sims D, et al. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet*. 2014;15(2):121–132.
57. Fouhy F, et al. The effects of freezing on faecal microbiota as determined using MiSeq sequencing and culture-based investigations. *PLoS One*. 2015;10(3):e0119355.
58. Kazantseva J, et al. Optimisation of sample storage and DNA extraction for human gut microbiota studies. *BMC Microbiol*. 2021;21(1):158.
59. Sinha R, et al. Assessment of variation in microbial community amplicon sequencing by the microbiome quality control (MBQC) project consortium. *Nat Biotechnol*. 2017;35(11):1077–1086.
60. Palkova L, et al. Evaluation of 16S rRNA primer sets for characterisation of microbiota in paediatric patients with autism spectrum disorder. *Sci Rep*. 2021;11(1):6781.
61. Darwish N, et al. Choice of 16S ribosomal RNA primers affects the microbiome analysis in chicken ceca. *Sci Rep*. 2021;11(1):11848.
62. Nearing JT, et al. Microbiome differential abundance methods produce different results across 38 datasets. *Nat Commun*. 2022;13(1):342.
63. Schloss PD. Identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in microbiome research. *mBio*. 2018;9(3):e00525–18.
64. Lynch SV, et al. Translating the gut microbiome: ready for the clinic? *Nat Rev Gastroenterol Hepatol*. 2019;16(11):656–661.
65. Food and Drug Administration. Database for Reference Grade Microbial Sequences (FDA-ARGOS). https://www.fda.gov/medical-devices/science-and-research-medical-devices/database-reference-grade-microbial-sequences-fda-argos. Updated December 26, 2019. Accessed November 29, 2021.
66. Mirzayi C, et al. Reporting guidelines for human microbiome research: the STORMS checklist. *Nat Med*. 2021;27(11):1885–1892.
67. STORMS. Strengthening The Organization and Reporting of Microbiome Studies. https://www.stormsmicrobiome.org. Accessed November 29, 2021.
68. Yatsunenko T, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486(7402):222–227.
69. Park J, et al. Shifts in the skin bacterial and fungal communities of healthy children transitioning through puberty. *J Invest Dermatol*. 2021;142(1):212–219.
70. Costello EK, et al. Bacterial community variation in human body habitats across space and time. *Science*. 2009;326(5960):1694–1697.
71. Kune GA. The Melbourne Colorectal Cancer Study: reflections on a 30-year experience. *Med J Aust*. 2010;193(11-12):648–652.
72. Suez J, et al. Artificial sweeteners induce glucose intolerance by altering the gut microbiota. *Nature*. 2014;514(7521):181–186.
73. Maier L, et al. Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature*. 2018;555(7698):623–628.
74. Zhang J, et al. Oral antibiotic use and risk of colorectal cancer in the United Kingdom, 1989-2012: a matched case-control study. *Gut*. 2019;68(11):1971–1978.
75. Rothschild D, et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature*. 2018;555(7695):210–215.
76. Jiang Z-D, et al. Genetic susceptibility to enteroaggregative Escherichia coli diarrhea: polymorphism in the interleukin-8 promotor region. *J Infect Dis*. 2003;188(4):506–511.
77. Balachandran VP, et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature*. 2017;551(7681):512–516.
78. Pleguezuelos-Manzano C, et al. Mutational signature in colorectal cancer caused by genotoxic pks^+ E. coli. *Nature*. 2020;580(7802):269–273.
79. Vermeulen R, et al. The exposome and health: where chemistry meets biology. *Science*. 2020;367(6476):392–396.
80. Walter J, et al. Establishing or exaggerating causality for the gut microbiome: lessons from

human microbiota-associated rodents. *Cell*. 2020;180(2):221–232.
81. Haile M, et al. CDC Supports Microbiome Science to Advance Infection Prevention, Clinical Care, and Public Health. https://blogs.cdc.gov/safehealthcare/cdc-supports-microbiome-science-to-advance-infection-prevention-clinical-care-and-public-health/. Updated June 25, 2021. Accessed November 29, 2021.
82. Zmora N, et al. Personalized gut mucosal colonization resistance to empiric probiotics is associated with unique host and microbiome features. *Cell*. 2018;174(6):1388–1405.
83. Wu GD, et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science*. 2011;334(6052):105–108.
84. O'Keefe SJD, et al. Fat, fibre and cancer risk in African Americans and rural Africans. *Nat Commun*. 2015;6(1):6342.
85. Zeevi D, et al. Personalized nutrition by prediction of glycemic responses. *Cell*. 2015;163(5):1079–1094.
86. Chen RY, et al. A microbiota-directed food intervention for undernourished children. *N Engl J Med*. 2021;384(16):1517–1528.
87. Byrd AL, et al. The human skin microbiome. *Nat Rev Microbiol*. 2018;16(3):143–155.
88. Flowers L, Grice EA. The skin microbiota: balancing risk and reward. *Cell Host Microbe*. 2020;28(2):190–200.
89. Sprockett D, et al. Role of priority effects in the early-life assembly of the gut microbiota. *Nat Rev Gastroenterol Hepatol*. 2018;15(4):197–205.
90. Dominguez-Bello MG, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A*. 2010;107(26):11971–11975.
91. Yassour M, et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci Transl Med*. 2016;8(343):343ra81.
92. Lynch SV, Vercelli D. Microbiota, epigenetics, and trained immunity. Convergent drivers and mediators of the asthma trajectory from pregnancy to childhood. *Am J Respir Crit Care Med*. 2021;203(7):802–808.
93. Fujimura KE, et al. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nat Med*. 2016;22(10):1187–1191.
94. Snitkin ES, et al. Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. *Sci Transl Med*. 2012;4(148):148ra16.
95. Johnson RC, et al. Investigation of a cluster of Sphingomonas koreensis infections. *N Engl J Med*. 2018;379(26):2529–2539.
96. Lieberman TD, et al. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat Genet*. 2011;43(12):1275–1280.
97. Wirbel J, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med*. 2019;25(4):679–689.
98. Thomas AM, et al. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med*. 2019;25(4):667–678.
99. Poore GD, et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature*. 2020;579(7800):567–574.
100. Sears CL, Salzberg SL. Microbial diagnostics for cancer: a step forward but not prime time yet. *Cancer Cell*. 2020;37(5):625–627.
101. Wilson MR, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N Engl J Med*. 2014;370(25):2408–2417.
102. Simner PJ, et al. Development and optimization of metagenomic next-generation sequencing methods for cerebrospinal fluid diagnostics. *J Clin Microbiol*. 2018;56(9):e00472–18.
103. Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet*. 2019;20(6):341–355.
104. Zhu N, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382(8):727–733.
105. Wright WF, et al. Progress report: Next-generation sequencing (NGS), multiplex polymerase chain reaction (PCR), and broad-range molecular assays as diagnostic tools for fever of unknown origin (FUO) investigations in adults [published online February 19, 2021]. *Clin Infect Dis*. https://doi.org/10.1093/cid/ciab155.
106. Damhorst GL, et al. Current capabilities of gut microbiome-based diagnostics and the promise of clinical application. *J Infect Dis*. 2021;223(12 suppl 2):S270–S275.
107. Wargo JA. Modulating gut microbes. *Science*. 2020;369(6509):1302–1303.
108. Gerardin Y, et al. Beyond fecal microbiota transplantation: developing drugs from the microbiome. *J Infect Dis*. 2021;223(suppl 3):S276–S282.
109. Paramsothy S, et al. Multidonor intensive faecal microbiota transplantation for active ulcerative colitis: a randomised placebo-controlled trial. *Lancet*. 2017;389(10075):1218–1228.
110. Narula N, et al. Systematic review and meta-analysis: fecal microbiota transplantation for treatment of active ulcerative colitis. *Inflamm Bowel Dis*. 2017;23(10):1702–1709.
111. Baruch EN, et al. Fecal microbiota transplant promotes response in immunotherapy-refractory melanoma patients. *Science*. 2021;371(6529):602–609.
112. Davar D, et al. Fecal microbiota transplant overcomes resistance to anti-PD-1 therapy in melanoma patients. *Science*. 2021;371(6529):595–602.
113. Zhang Y, et al. Excretion of SARS-CoV-2 through faecal specimens. *Emerg Microbes Infect*. 2020;9(1):2501–2508.
114. Zhang W, et al. Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. *Emerg Microbes Infect*. 2020;9(1):386–389.
115. Drewes JL, et al. Transmission and clearance of potential procarcinogenic bacteria during fecal microbiota transplantation for recurrent Clostridioides difficile. *JCI Insight*. 2019;4(19):e130848.
116. Wilcox MH, et al. The efficacy and safety of fecal microbiota transplant for recurrent *Clostridium difficile* infection: current understanding and gap analysis. *Open Forum Infect Dis*. 2020;7(5):ofaa114.
117. DeFilipp Z, et al. Drug-resistant *E. coli* bacteremia transmitted by fecal microbiota transplant. *N Engl J Med*. 2019;381(21):2043–2050.
118. Food and Drug Administration. Fecal Microbiota for Transplantation: New Safety Information - Regarding Additional Protections for Screening Donors for COVID-19 and Exposure to SARS-CoV-2 and Testing for SARS-CoV-2. https://www.fda.gov/safety/medical-product-safety-information/fecal-microbiota-transplantation-new-safety-information-regarding-additional-protections-screening. Updated April 9, 2020. Accessed November 29, 2021.
119. Carlson PE. Regulatory considerations for fecal microbiota transplantation products. *Cell Host Microbe*. 2020;27(2):173–175.
120. Tariq R, et al. Low cure rates in controlled trials of fecal microbiota transplantation for recurrent clostridium difficile infection: a systematic review and meta-analysis. *Clin Infect Dis*. 2019;68(8):1351–1358.
121. Williams S. Making poop profitable. *Scientist*. 2021;35(3):46–49.
122. McGovern BH, et al. SER-109, an investigational microbiome drug to reduce recurrence after Clostridioides difficile infection: lessons learned from a phase 2 trial. *Clin Infect Dis*. 2021;72(12):2132–2140.
123. Blount KF, et al. Restoration of bacterial microbiome composition and diversity among treatment responders in a phase 2 trial of RBX2660: an investigational microbiome restoration therapeutic. *Open Forum Infect Dis*. 2019;6(4):ofz095.
124. Kwak S, et al. Impact of investigational microbiota therapeutic RBX2660 on the gut microbiome and resistome revealed by a placebo-controlled clinical trial. *Microbiome*. 2020;8(1):125.
125. Feuerstadt P, et al. SER-109, an oral microbiome therapy for recurrent *Clostridioides difficile* infection. *N Engl J Med*. 2022;386(3):220–229.
126. Hourigan SK, et al. Microbiome changes associated with sustained eradication of Clostridium difficile after single faecal microbiota transplantation in children with and without inflammatory bowel disease. *Aliment Pharmacol Ther*. 2015;42(6):741–752.
127. Chen LA, Sears C. Prebiotics, probiotics, and synbiotics. In: Bennett J, et al., eds. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*. Elsevier; 2020:19–25.
128. Freedman SB, et al. Multicenter trial of a combination probiotic for children with gastroenteritis. *N Engl J Med*. 2018;379(21):2015–2026.
129. Schnadower D, et al. Lactobacillus rhamnosus GG versus placebo for acute gastroenteritis in children. *N Engl J Med*. 2018;379(21):2002–2014.
130. Piewngam P, et al. Pathogen elimination by probiotic Bacillus via signalling interference. *Nature*. 2018;562(7728):532–537.
131. Lam KN, et al. Precision medicine goes microscopic: engineering the microbiome to improve drug outcomes. *Cell Host Microbe*. 2019;26(1):22–34.
132. Dedrick RM, et al. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus. *Nat Med*. 2019;25(5):730–733.
133. Schooley RT, et al. Development and use of personalized bacteriophage-based therapeutic

cocktails to treat a patient with a disseminated resistant Acinetobacter baumannii infection. *Antimicrob Agents Chemother*. 2017;61(10):e00954–17.
134. Kurtz CB, et al. An engineered *E. coli* Nissle improves hyperammonemia and survival in mice and shows dose-dependent exposure in healthy humans. *Sci Transl Med*. 2019;11(475):eaau7975.
135. Quigley J, et al. Novel antimicrobials from uncultured bacteria acting against Mycobacterium tuberculosis. *mBio*. 2020;11(4):e01516–20.
136. Zhang F, et al. A marine microbiome antifungal targets urgent-threat drug-resistant fungi. *Science*. 2020;370(6519):974–978.
137. Lehtinen I. *Comparison of Normalization and Statistical Testing Methods of 16S rRNA Gene Sequencing Data*. Master's thesis. Department of Computer Science, Aalto University, Espoo, Finland; 2018:67.
138. Evans AS. Causation and disease: the Henle-Koch postulates revisited. *Yale J Biol Med*. 1976;49(2):175–195.
139. Falkow S. Molecular Koch's postulates applied to microbial pathogenicity. *Rev Infect Dis*. 1988;10(suppl 2):S274–S276.
140. Zymo Research. 16S Sequencing vs Shotgun Metagenomic Sequencing. https://www.zymoresearch.com/blogs/blog/16s-sequencing-vs-shotgun-metagenomic-sequencing. Accessed November 13, 2021.
141. Truong DT, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods*. 2015;12(10):902–903.
142. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*. 2014;15(3):R46.
143. Altschul SF, et al. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–410.
144. Huson DH, et al. MEGAN analysis of metagenomic data. *Genome Res*. 2007;17(3):377–386.
145. Albanese D, et al. MICCA: a complete and accurate software for taxonomic profiling of metagenomic data. *Sci Rep*. 2015;5:9743.
146. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J*. 1948;27(3):379–423.
147. Simpson EH. Measurement of diversity. *Nature*. 1949;163(4148):688.
148. Del Chierico F, et al. Choice of next-generation sequencing pipelines. In: Mengoni A, et al., eds. *Bacterial Pangenomics: Methods in Molecular Biology*. Humana Press; 2015(1231). https://doi.org/10.1007/978-1-4939-1720-4_3.