Annual Review of Biomedical Data Science

Metatranscriptomics for the Human Microbiome and Microbial Community Functional Profiling

Yancong Zhang,[1,2,*] Kelsey N. Thompson,[1,2,*] Tobyn Branck,[1,2,3] Yan Yan,[1,2] Long H. Nguyen,[1,4,5] Eric A. Franzosa,[1,2,†] and Curtis Huttenhower[1,2,6,†]

[1]Harvard Chan Microbiome in Public Health Center and Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts 02115, USA; email: franzosa@hsph.harvard.edu, chuttenh@hsph.harvard.edu

[2]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA

[3]Department of Systems, Synthetic, and Quantitative Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

[4]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA

[5]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02108, USA

[6]Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, Massachusetts 02115, USA

Annu. Rev. Biomed. Data Sci. 2021. 4:279–311

First published as a Review in Advance on May 13, 2021

The *Annual Review of Biomedical Data Science* is online at biodatasci.annualreviews.org

https://doi.org/10.1146/annurev-biodatasci-031121-103035

*These authors contributed equally to this article

†These authors jointly supervised this article

Keywords

metatranscriptome, meta-omics, functional activity, RNA-seq

Abstract

Shotgun metatranscriptomics (MTX) is an increasingly practical way to survey microbial community gene function and regulation at scale. This review begins by summarizing the motivations for community transcriptomics and the history of the field. We then explore the principles, best practices, and challenges of contemporary MTX workflows: beginning with laboratory methods for isolation and sequencing of community RNA, followed by informatics methods for quantifying RNA features, and finally statistical methods for detecting differential expression in a community context. In the

second half of the review, we survey important biological findings from the MTX literature, drawing examples from the human microbiome, other (nonhuman) host-associated microbiomes, and the environment. Across these examples, MTX methods prove invaluable for probing microbe–microbe and host–microbe interactions, the dynamics of energy harvest and chemical cycling, and responses to environmental stresses. We conclude with a review of open challenges in the MTX field, including making assays and analyses more robust, accessible, and adaptable to new technologies; deciphering roles for millions of uncharacterized microbial transcripts; and solving applied problems such as biomarker discovery and development of microbial therapeutics.

MICROBIAL COMMUNITY TRANSCRIPTOMICS

The degree to which sequencing has become a dominant tool for biology in the mere two decades since the Human Genome Project is astonishing. It is thus easy to forget that genomics was launched in microbes, with the genome of *Haemophilus influenzae* completed just a few years prior (1). However, it has become cliché to note that decoding a genome does not reveal its functions or phenotypes—and microbes are also to thank for the first high-throughput, whole-genome functional assays in the form of gene expression profiling (2). Transcriptomics—first by microarrays and now predominantly by RNA sequencing (RNA-seq)—provides one of the most tractable links between the genetic potential of an organism and its molecular activity. Recently, just as DNA sequencing has moved from single genomes to microbial community shotgun metagenomes [metagenomics (MGX)] (3), the same has become true for the use of metatranscriptomics (MTX) to profile whole-community RNA. While both technical and analytical protocols remain challenging (4), MTX is a crucial tool for understanding microbial behavior in situ since microbes can express completely different biochemical programs in isolate culture than in their natural community settings (5). The latter is often crucial for human health (6–9) and the environment (10), and in this review we thus summarize the history of microbial community transcriptomics, the current state of methods for and discoveries from MTX, and potential gaps and next steps in the field.

In the human microbiome, MTX has the potential to provide both basic biological knowledge and epidemiological biomarkers, i.e., high-dimensional diagnostics for current health status or prognostics for future health outcomes (6, 11). MTX biomarkers occupy a unique niche in molecular epidemiology (12): Unlike human genetics, microbiome gene expression changes over time; MTX changes much more rapidly than microbiome membership (MGX), thus providing a shorter memory of more recent exposures; due to the microbiome's personalization (13), MTX can be much more subject specific than human transcriptomes; like tissue-specific transcription, MTX is specific to body site, capturing local molecular activity; and in limited cases, human and microbial transcriptional profiles can be directly combined to observe host–microbe interactions (14). While microbiome diversity and membership have been intensively explored via MGX for research and commercialization (3, 15, 16), as have human transcriptional biomarkers (17, 18), the application of MTX to human population health remains in its infancy.

MTX for more general mechanistic characterization also remains challenging, but it has been substantially better explored, particularly in non-human-associated and environmental microbial communities (19, 20). While MGX is usually most sensitive to bacteria, it also accesses archaeal, eukaryotic (e.g., fungal), and DNA viral community members; MTX further adds some visibility into RNA viruses [of particular interest in the age of COVID-19 (coronavirus disease 2019)] (9). Regardless of kingdom, laboratory isolates and microbes existing in complex communities exhibit vastly different growth profiles, transcription, and physiology (21, 22). For example, each laboratory-based *Escherichia coli* cell might harbor 1,300–1,800 mRNA molecules, in contrast to

only a few hundred transcripts per cell in bacterioplankton communities in situ (20). This is in addition to differences in transcriptional programs triggered by interspecies microbial interactions, nutrient availability (either globally or in local microregions), other local chemical differences (e.g., oxygen or pH), biofilm structure, and growth rate (23). Since these regulatory programs are at best extremely difficult to mimic in a laboratory setting, MTX is the only current means by which they can be explored—with the added benefit of improving our understanding of other basic functions of RNA molecules in microbial communities, such as determining the transcriptional architecture of genes, their start sites, and 5′ and 3′ ends, as well as identifying noncoding RNAs, including antisense and small noncoding RNAs (24, 25). Finally, the emergent phenotypes of non-human-associated microbial communities are often themselves nearly as important as human health, be they in bioreactors for biofuel production, bioremediation, livestock production, or climate change (26–31).

Despite this potential, MTX has been largely underutilized to date, for both experimental and analytical reasons. Extraction protocols able to retrieve fragile RNA molecules without bias from a wide range of diverse microbial cell structures can be extremely sensitive (32). Once extracted, community RNA can be easily degraded, not only by the extraction procedure itself but also by nucleases that are natively abundant within communities. MTX is more susceptible than MGX to the often complex chemical makeup of the matrix surrounding microbial cells in a community, and even if extracted successfully, metatranscriptomes must be profiled at substantially greater depth in most settings to avoid underdetection (12). There are few validated computational methods available for the bioinformatics of raw MTX sequence handling—that is, the process of transforming reads into quantified assemblies or transcriptional or taxonomic profiles (5). Likewise, downstream statistics are complicated by zero inflation and dynamic range, akin to the combination of the hurdles encountered by MGX and single-cell RNA-seq (scRNA-seq) (33). This leaves striking knowledge gaps in the field: At least two-thirds of microbial genes in the human microbiome, and often over 90% in non-human-associated communities, remain uncharacterized (34). To our knowledge, no MTX biomarkers have reached even preclinical testing to date. The field is thus rich in potential, and in this review, we cover the currently available laboratory and analysis protocols; statistical approaches appropriate for MTX; examples of successful MTX studies in different human, host-associated, and environmental ecosystems; and possible next steps in applying and extending MTX-based biological knowledge.

METHODS FOR METATRANSCRIPTOMICS

Isolation and Processing of Microbiome mRNA

As with individual organisms, microarrays were the first platform utilized to determine the gene expression profiles of microbial communities at scale (35). Such approaches have been used not only for transcript abundance measurement but also to detect the boundaries of untranslated regions (36–38), to identify noncoding RNA (39, 40), to detect small peptides (41), and to address a variety of context-specific biological questions [e.g., aquatic biofilm responses to xenobiotics (42)]. Even more than for individual organisms, the design of the probes used in each microarray is critical for the quality of the resulting data in order to avoid ambiguity, cross-hybridization, or bias. Typically, community arrays have been designed to probe the dominant known taxa in a community of interest by adapting tools for isolate transcriptomics (43). In the absence of reference sequences, this often has relied on cDNA (complementary DNA) libraries (42), but these were successful in community differential expression, e.g., by environmental parameters such as oxygen availability or salinity (44). For many communities, microarrays still offer sufficient probe density for dense tiling, deal well with low-biomass samples, and avoid the need to deplete ribosomal or nonmicrobial RNA, as encountered during shotgun sequencing.

The extreme diversity of microbial communities, however, and (especially until recently) their lack of reference isolates make shotgun MTX via RNA-seq a more practical approach in most settings. Again paralleling single organisms, the earliest MTX used low-throughput sequencing of cDNA clones to characterize expression profiles (45, 46); these studies of soil and marine communities also had the benefit of relatively low-complexity ecologies, little or no nonmicrobial contamination, and extremely large sample volume. Pyrosequencing and other high-throughput technologies have made community MTX even more efficient since per-sample pan-genomes encompassing millions of transcripts can easily require gigabases of sequencing to saturate adequately. The 454 platform was used to profile community gene expression in both soil and ocean water (47–51), finding abundant transcripts for ammonia cycling, carbon fixation (i.e., photosynthesis), cellular maintenance, and archaeal contributions—along with many, many novel transcripts. More recently, studies have used Illumina sequencing platforms to characterize MTX in diverse host- and non-host-associated (free-living) communities (6, 9, 12, 30, 52–54), as discussed below. While long-read (e.g., PacBio or Nanopore) and linked-read [e.g., 10×, Hi-C (55)] technologies are now regularly applied to MGX (particularly amplicons), their use for MTX [e.g., surveillance for plant pathogens, including RNA viruses (56)] remains limited, albeit with substantial potential when cost effective.

The broad strokes of MTX data generation and analysis resemble those from a combination of MGX and single-organism transcriptomics (**Figure 1**). RNA extraction from microbial communities can be carried out by a variety of commercially available kits (e.g., Qiagen's PowerMicrobiome or AllPrep, Thermo Fisher's PureLink). Once sufficiently high–molecular weight RNA is extracted, it must be cleaned (separated from DNA and other biomolecules) while evenly retaining as much biomass as possible [typically by SPRI (solid phase reversible immobilization) beads (10)]. Next, noncoding RNAs are generally removed prior to cDNA library construction, with major methods including nuclease targeting and hybrid capture (detailed below) (57). Often, RNA (cDNA) and DNA (MGX) from the same community will be jointly extracted and sequenced in parallel, allowing for more flexible and accurate analyses later on (also detailed below). Once reverse-transcribed, MTX libraries are sequenced essentially identically to MGX, but with both similarities and important differences in their subsequent analysis. Quality control (QC) may be similar, as are reference-based mapping approaches, but isolate, MGX, and MTX assembly differ substantially, for example. Finally, once transcript models are defined and quantified, statistics for analysis must incorporate aspects of traditional RNA-seq (e.g., read normalization), MGX (DNA copy number variation), and scRNA-seq (zero inflation and extreme dynamic range) (58).

Perhaps the defining technical challenge of MTX is the ability to obtain a sufficient quantity of high-quality, intact mRNA (nonribosomal) that evenly represents the community of interest. Experimental strategies have been developed to tackle the short half-lives of mRNA (59) and high amounts of inhibitory substances coextracted, particularly in mixed communities (60). By comparing three different commercial RNA extraction kits (61), Reck et al. (62) studied the stability of the metatranscriptome in stool samples under various preservation and storage conditions. At the time, the MoBio (now Qiagen) PowerMicrobiome kit provided the best RNA yield and purity. RNAlater efficiently preserved mRNA integrity over time even at room temperature but introduced a small bias in the transcriptional profile at the beginning that was subsequently maintained throughout storage (9), while RNAprotect introduced a smaller bias but showed substantial mRNA decay when storage times exceeded 24 h. While these issues are likely not unique to the chemistry of MTX relative to isolates, they are uniquely important in a community context, when samples must be collected in the wild and are not necessarily immediately available in a laboratory under controlled conditions.

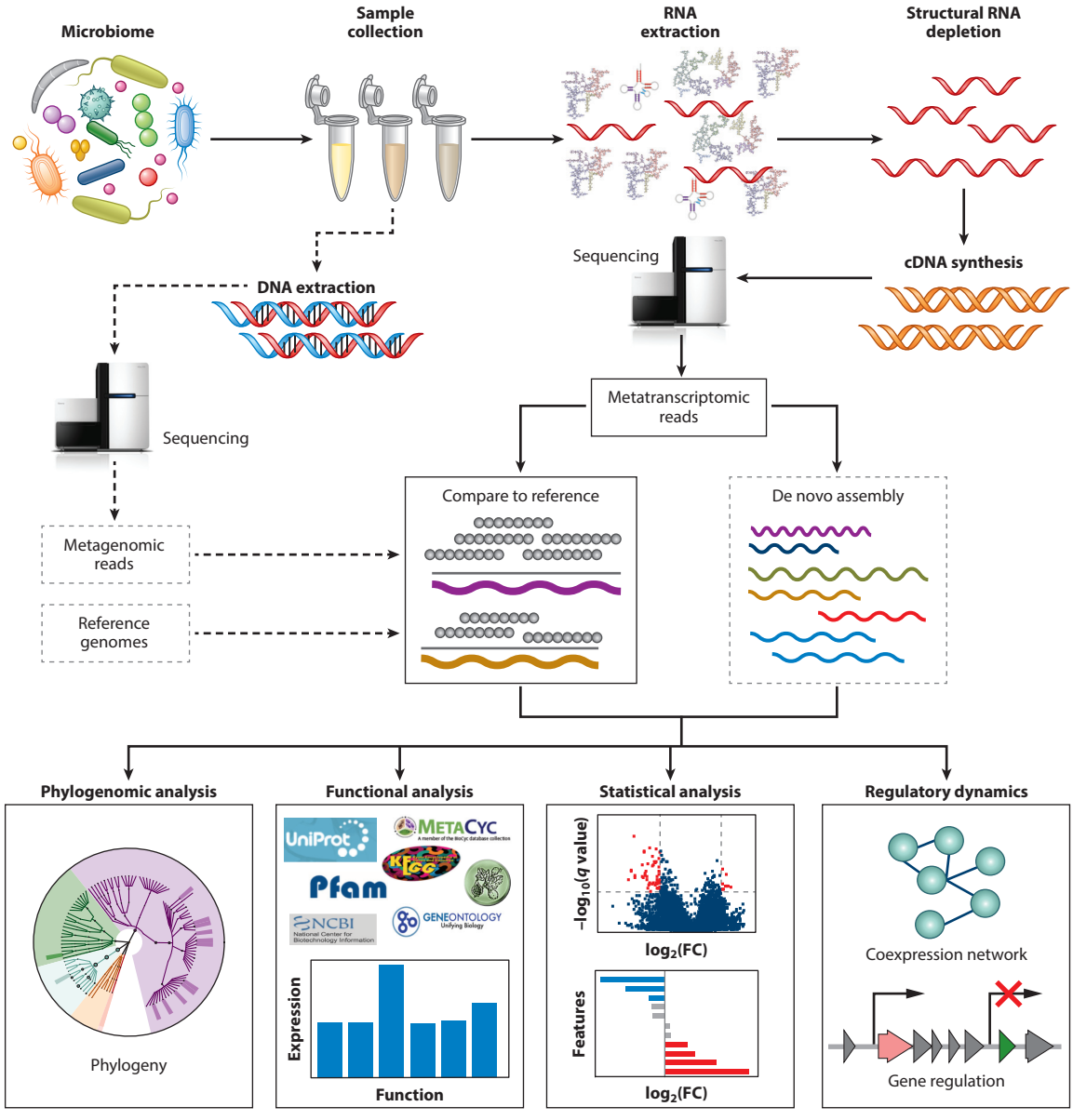

Figure 1

MTX laboratory protocols, sequencing, and analysis. Community MTX, like single-organism transcriptomics, begins with RNA extraction and reverse-transcription to cDNA for library preparation. This can be complicated by the chemical diversity of microbial communities, RNA instability, biomass availability, and the predominance of non-protein-coding or nonmicrobial RNAs. A corresponding metagenome can be processed in parallel and combined with reference sequences for MTX analysis. Common steps include quality control of the sequencing reads, alignment to reference sequences or de novo assembly, taxonomic and functional annotation, statistical analysis (e.g., differential expression), and regulatory analysis (e.g., dynamics and coexpression). Abbreviations: cDNA, complementary DNA; FC, fold-change.

The challenge of separating non-protein-coding RNAs from mRNAs in microbial communities is perhaps just as great a hurdle. The fraction of ribosomal RNA (rRNA) to transfer RNA (tRNA) accounts for more than 95% of total cellular RNA in prokaryotes (63); thus, while it is possible to just deep-sequence a complete cDNA library, it is rarely cost effective. Unlike eukaryotes, the majority of prokaryotic mRNA lacks a poly-A tail, and thus the relatively simple option of oligo-dT selection is not available. Various approaches have been developed to enrich for the desired mRNA fraction (32, 63–65), including subtractive hybridization (Hyb) with rRNA-specific probes (66) [e.g., MICROBExpress from Ambion, RiboMinus from Thermo Fisher, Ribo-Zero from Illumina, other approaches using noncommercial and sample-specific antirRNA probes (67)] and exonuclease digestion preferentially acting on rRNA (e.g., mRNA-ONLY from Epicentre) (25, 49). He et al. (64) validated the performance of Hyb and exonuclease digestion using Illumina-based RNA-seq of synthetic microbial metatranscriptomes. Their results suggested that the Hyb method introduced less bias in the relative proportion of the mRNA population compared to exonuclease digestion. Interestingly, however, host RNA contamination (e.g., from human microbiomes) can remain a problem in this setting, since eukaryotic rRNA is not removed by bacterial rRNA probes. Depletion efficiencies can also vary greatly among samples and sample types, and extremely high efficiencies are still needed to lower sequencing costs. With uncertainty surrounding the availability of the widely used Illumina Ribo-Zero kit for the past few years, this is a surprisingly narrow but crucial area in which further technical developments could greatly benefit MTX.

Relatedly, while MTX analysis is discussed in detail below, perhaps its most important and overlooked feature is appropriately combining it with MGX paired with each sample. Biologically, the argument is simple: MGX represents the genetic potential of a community and MTX its expressed functional activity. However, MGX also represents a change in the underlying DNA copy number of each RNA transcript, independently of its expression level. This is akin to copy number variation or aneuploidy in single-organism transcriptomics (68), but whereas such events are typically minor in single organisms, they are universal in microbial communities: The replication or death of each microbial cell changes relative DNA copy numbers. Thus, while MTX can be analyzed in isolation (25, 69, 70), it is arguably more appropriate to jointly analyze community DNA and cDNA from the same sample, allowing the integration of transcript abundances with corresponding gene copy numbers from the community (9, 30, 71, 72). Strikingly, bioinformatics or statistics for this process are not yet standardized and can produce substantially different results, as discussed below. Likewise, while long-read technologies provide a new and potentially highly informative way to assay (near) full-length mRNAs from communities (68, 73), no methods have yet been developed to take advantage of this unique measurement type (74).

Metatranscriptomic Data Informatics

Typical MTX bioinformatics include at least four main steps (**Figure 1**): (*a*) QC of shotgun sequencing to remove or trim spurious/erroneous reads, (*b*) alignment to reference sequences or de novo assembly to profile transcript abundances, (*c*) functional and taxonomic characterization to identify active elements and members of the community, and (*d*) statistical analyses to normalize expression and identify changes between different conditions (e.g., differential expression or coexpression dynamics). Several efficient workflows have been developed to address combinations of these, such as the pipeline developed by Leimena et al. (75), MetaTrans (76), IMP (77), COMAN (78), SAMSA2 (79), HUMAnN (80), and SqueezeMeta (81) (see the sidebar titled Software Considerations for Metatranscriptome Analysis). Interestingly, while subsets of these methods have been recently reviewed elsewhere (82, 83), their performance has in general not been comparatively evaluated. Thus, here we focus on the conceptual needs and challenges of performing these main steps.

SOFTWARE CONSIDERATIONS FOR METATRANSCRIPTOME ANALYSIS

Although there are to date very few computational workflows specific for MTX, several approaches can be used or adapted, many of which have been recently reviewed elsewhere (82, 83). Below we list some important considerations in choosing an analysis approach and software implementation.

Taxonomy Versus Function

Quantifying MTX features in a community is a key task of MTX analysis. Features can include functional units (e.g., transcripts or pathways) or the taxa transcribing them. For example, the pipelines developed by Leimena et al. (75), MetaTrans (76), COMAN (78), SAMSA2 (79), HUMAnN (80), IMP (77), and SqueezeMeta (81) profile both taxa and functions, while FMAP (84) is an example of a dedicated functional profiling system.

Assembly Versus Reference

Homology-based (i.e., mapping) analysis is the most common approach to profile MTX features with high sensitivity, and it is the backbone of many pipelines (e.g., Leimena et al., MetaTrans, COMAN, FMAP, SAMSA2, HUMAnN). Several other MTX pipelines integrate optional MTX assembly steps that allow for novel transcript discovery, including IMP, SqueezeMeta, and MOSCA (**https://github.com/iquasere/MOSCA**).

How to Handle Paired DNA

When metagenomes (DNA) derived from the same samples are available, paired MGX and MTX can be analyzed jointly, allowing MGX to profile the taxonomy of a community and MTX features to be assigned to the resulting MGX-based taxa. Pipelines like HUMAnN, FMAP, IMP, and SqueezeMeta support processing of paired MGX and MTX studies.

Quality Control of Raw Reads

Some pipelines support QC steps that are somewhat unique to MTX, such as handling rRNA and structural RNA contamination in addition to simpler low-quality reads (e.g., MetaTrans, COMAN, SAMSA2, SqueezeMeta) or differentiating microbial- versus host-derived reads (e.g., FMAP, IMP).

How to Handle Read Pairing

Paired reads are sometimes used for assembling metatranscriptomes in MTX-specific assemblers [e.g., IDBA-MT (85), IDBA-MTP (86), TAG (87)] or in some assembly-based pipelines (e.g., IMP, SqueezeMeta).

Long-Read Support

Pipelines such as SqueezeMeta have started to support long-read processing (e.g., Pacific Biosciences or Oxford Nanopore).

Visualization Support

Several pipelines provide utilities specifically to visualize the results of MTX analysis, such as HUMAnN, MetaTrans, and SAMSA2.

Differential Expression Support

While differential expression of MTX features is often analyzed statistically using existing (sc)RNA-seq approaches, several pipelines also encourage tests tailored to associate alterations in MTX with sample phenotypes. Often this integrates tools designed for single-organism RNA-seq [e.g., DESeq2 (88), edgeR (89)] for identifying differential expression, including MetaTrans, COMAN, FMAP, and SAMSA2.

Quality control. MTX generally requires greater depth of sequencing than does an MGX profile of the same community, which is more an aspect of experimental design than QC per se. Conceptually, if the rarest organism is 100 times less abundant than the most common organism, then the ~1% of the DNA reads that it recruits must be sufficient to detect it metagenomically. For microbial genomes of (very approximately) 5 Mnt (mega-nucleotides), this makes a minimum of 1 Gnt (giga-nucleotides) DNA sufficient per sample (3). If transcripts are themselves regulated over a >100-fold range, however, then, in principle, 10,000-fold deeper sequencing would be required to reliably detect both the rarest and the most abundant transcripts, which is clearly infeasible. MTX designs must thus be prepared for technical zeros (nondetects), while also acknowledging that in many cases, there may be little prior knowledge about the diversity of the underlying community, the relative abundances of each species, genome sizes, and the most actively transcribed genes. As a result—akin to some proteomics protocols (90)—most MTX protocols are biased toward the most abundant features. It is thus also not clear what sequencing depth is appropriate for MTX in general, without fairly extensively specifying the desired analysis goals and estimated ecological parameters. A reasonable rule of thumb for MTX sequencing depth is that ~10× that of accompanying MGX is plausible, but that will still only reach saturation of the most abundant transcripts (12).

After an MTX cDNA library is sequenced, much of its initial QC parallels that of MGX or single-organism genomics/transcriptomics. This includes removal of sequencing adapters, trimming low-quality bases, and removing potential low-quality or contaminant reads. Like MGX, sequencing data from host-associated communities can contain a high percentage of host reads to be depleted. Similarly, despite the rRNA depletion protocols described above, noncoding RNAs can often still dominate after sequencing and require computational depletion. Multiple tools are appropriate from genomics and MGX to handle individual steps, such as FastQC (**https://www.bioinformatics.babraham.ac.uk/projects/fastqc/**), FaQCs (91), Trimmomatic (92), and FASTX (**http://hannonlab.cshl.edu/fastx_toolkit/**) for controlling reads quality and SortMeRNA (93) and barrnap (94) for identifying and removing rRNAs. KneadData (**http://huttenhower.sph.harvard.edu/kneaddata**) integrates all of these QC steps for MGX or MTX data.

Mapping-based metatranscriptomic analyses. QCed MTX reads can subsequently be analyzed in a manner similar to single-organism transcriptomics either by mapping to reference genomes or by assembly into transcripts. The former is more commonly used for communities, mainly due to being far less computationally intensive and less well suited for the detection of low-abundance transcripts. MTX-specific transcript assemblers are also generally understudied to date (95). However, like MGX, mapping relies on the availability and accuracy of reference sequences and is challenging in communities with novel organisms, multiple closely related strains, or ambiguous orthologous or paralogous gene families and repetitive sequences (5). Many of the studies discussed below have used mapping-based strategies (6, 9, 12, 30, 72), and mapping-based approaches have also been integrated in several workflows, including that of Leimena et al. (75), MetaTrans (76), COMAN (78), FMAP (84), SAMSA2 (79), and HUMAnN (80). These methods generally query MTX reads against gene or genome taxonomic and functional reference databases using accelerated alignment tools, such as bowtie2 (96), BWA (97), DIAMOND (98), or MMseqs2 (99). After overcoming these issues, the resulting SAM (sequence alignment/map)/BAM (binary SAM) files can be converted into counts representing the occurrence of each gene/functional model in each organism (if available), length and depth normalized, and annotated using preexisting resources [e.g., NCBI's (National Center for Biotechnology Information's) RefSeq (100), Pfam domains (101), GO (Gene Ontology) terms (102), KEGG (Kyoto Encyclopedia of Genes and

Genomes; 103), COG (Clusters of Orthologous Groups; 104), or UniProt (105)] to characterize phylogenetic and functional assignments, much as in single-organism RNA-seq.

While the questions "Who's there?" and "What are they doing?" are often analyzed jointly in MGX, gene function and taxon of origin can be much more loosely coupled during MTX analysis. While MGX constructs (in principle) a complete view of community composition, including DNA that is dormant, cell-free, or nonviable, MTX inherently does not aim to profile the entire community. Conversely, if a transcript can be confidently assigned to a microbe, it is likely that microbe is present and viable. In general, it is thus still important to know which organisms are transcribing which functions and are therefore active, e.g., identification of which organism provides the bulk of oil degradation during spill remediation (10). This should not be confused with taxonomic profiling from MTX, however, since direct application of tools for MGX profiling to MTX can produce misleading results. MetaPhlAn (106), for example, calculates taxon abundance by aggregating multiple loci expected to occur at the same DNA copy number within each clade; this assumption does not hold in metatranscriptomes. The same biology precludes meaningful taxonomic profile calculation from short-read classifiers such as Kraken (107) or mOTUs2 (108). The most precise MTX studies have thus generally used the subset of data that can confidently be assigned to a gene family (even if uncharacterized) and a taxon (even if unknown) simultaneously, providing a view that resembles an aggregate of single-organism transcriptomes (80). Examples include studies that have combined MGX and MTX to profile taxonomic shifts during antibiotic therapy (109), studies that have identified nucleotide metabolism depleted in vaginal microbiomes of women of Adrican ancestry (110), and studies that have found profound transcriptional changes induced by the consumption of a Western diet in mice (111).

With or without knowledge of which organisms encode active transcripts, the fundamental task of MTX is to quantify gene family expression in a community, or what happens when conditions change. This entails functional profiling in which feature abundances—be they genes/transcripts, families such as EC (Enzyme Commission) numbers (112) or Pfam numbers (113), or pathways [e.g., GO (114), KEGG (115)]—are quantified from sequences. Mapping-based tools validated for this task in MTX include MetaCLADE (116), SAMSA2 (79), and HUMAnN (80). These methods all roughly search reads (either as nucleotides or as multiframe translations) against a reference set (of whole genomes, genes, or gene families/pan-genomes). Annotations of the resulting hits are then resolved, quantified [in units directly paralleling those of single-organism RNA-seq, e.g., copies per million (CPMs) (117)], and grouped based on annotations such as Pfam (101), GO (102), KEGG (103), COG (104), SEED subsystems (118), or UniRef families (119). Like MGX functional profiling, MTX mapping strategies are heavily reliant both on their algorithm and on the associated reference database(s) with which they are used. Reliance on homology-based functional annotation in MTX—of necessity—has thus particularly highlighted the abundance of uncharacterized and poorly characterized transcripts in most wild microbial communities (30).

Assembly-based metatranscriptomic analyses. Assembly-based community analysis approaches have the advantages of not requiring references and allow for the identification of novel expressed protein-coding genes or noncoding RNAs. This novelty comes at the cost of sensitivity, however, since only sufficiently covered sequences (DNA or RNA) can be contiguously assembled. When paired metagenomes are available, they can be assembled separately and MTX reads mapped to them, somewhat avoiding the need for more complex metatranscriptome assembly (120, 121). Although highly dependent on the depth of sequencing and complexity of the underlying community, this approach typically incorporates roughly half of the sequencing reads (16). When paired metagenomes are not available, QCed MTX reads can be assembled directly

into longer fragments using de novo assemblers. However, unlike single organisms (122–124), relatively few assemblers are designed specially for the metatranscriptome, such as IDBA-MT (85), IDBA-MTP (86), and TAG (87). In some cases (125), single-organism transcriptome assemblers such as Trinity (122) still outperformed Oases (126), Trans-Abyss (127), and IDBA-MT in MTX assembly. Bioinformatic workflows such as IMP (77), SqueezeMeta (81), and MOSCA (**https://github.com/iquasere/MOSCA**) have integrated optional MTX assembly steps with multiple assembly tools available to choose from. However, with few methods truly customized to the task or quantitatively evaluated for it, MTX analysis by community assembly is still in its early stage.

Epidemiology and Biostatistics for Metatranscriptomes

Having quantified functional activity from a collection of biosamples using any of the methods above, a common next step in MTX analysis is to associate changes in community activity with sample phenotypes or environmental parameters (**Figure 2**). Such analyses must also be performed carefully, as the numerical properties of MTX data can induce spurious associations if appropriate statistical methods are not followed. For example, similar to single-organism RNA-seq (128, 129), MTX measurements are based on integer counts of sequencing reads, which are sensitive to sample sequencing depth and transcript length (in addition to bioactivity). Transcript counts are commonly inflated with zero values (**Figure 2*a***), many of which correspond to nondetection events rather than an absence of expression. In addition, transcript abundance values within a species may span several orders of magnitude, as can the abundance of a single transcript across samples, thus violating the normality assumptions of many common statistical tests. Statistical methods for single-organism RNA-seq, including edgeR (89), CuffDiff 2 (130), DESeq2 (88), limma (131), and NOISeq (132), have been developed to address these challenges, and such methods are often packaged as components of dedicated MTX workflows (76, 78, 84).

Statistical analysis of MTX data poses two additional challenges relative to single-organism RNA-seq that both involve transcript–gene copy relationships. Single-organism RNA-seq generally assumes that each cell in each sample contains one copy of each gene (constant functional potential). This is often not the case in microbial community samples, where genes can be lost or duplicated within sample-specific strains (**Figure 2*a***), with gene loss representing an additional source of RNA zeroes (not attributable to canonical downregulation). More generally, as the number of genes encoding a transcript increases in a community (e.g., due to increased taxonomic abundance of species encoding the function), the transcript's abundance tends to increase in kind (**Figure 2*b***). At the community level, this phenomenon manifests as a strong correlation between functions' DNA- and RNA-level abundance estimates across samples (9). Hence, many perceived differences in MTX abundance are expected to arise from underlying variations in gene copy number (metagenomic abundance) rather than differential expression or functional activity.

To compensate for the strong baseline dependency of community functional activity on functional potential, one needs to estimate the relative expression of a function, i.e., the degree to which it is over- or underexpressed in a metatranscriptome relative to the abundance of community genes encoding the function (**Figure 2*c***). When community genes can be assigned to source species, one approach for estimating relative expression is to normalize transcript abundances within species [sometimes called taxon-specific scaling (133)]. This approach to MTX normalization assumes that all genes within a species have the same copy number (i.e., ignoring potential gene loss or duplication events), and it effectively transforms an MTX dataset into a collection of separate single-organism RNA-seq datasets. Alternatively, for microbial communities profiled with paired MTX and MGX sequencing, one can directly normalize each gene family's community RNA abundance

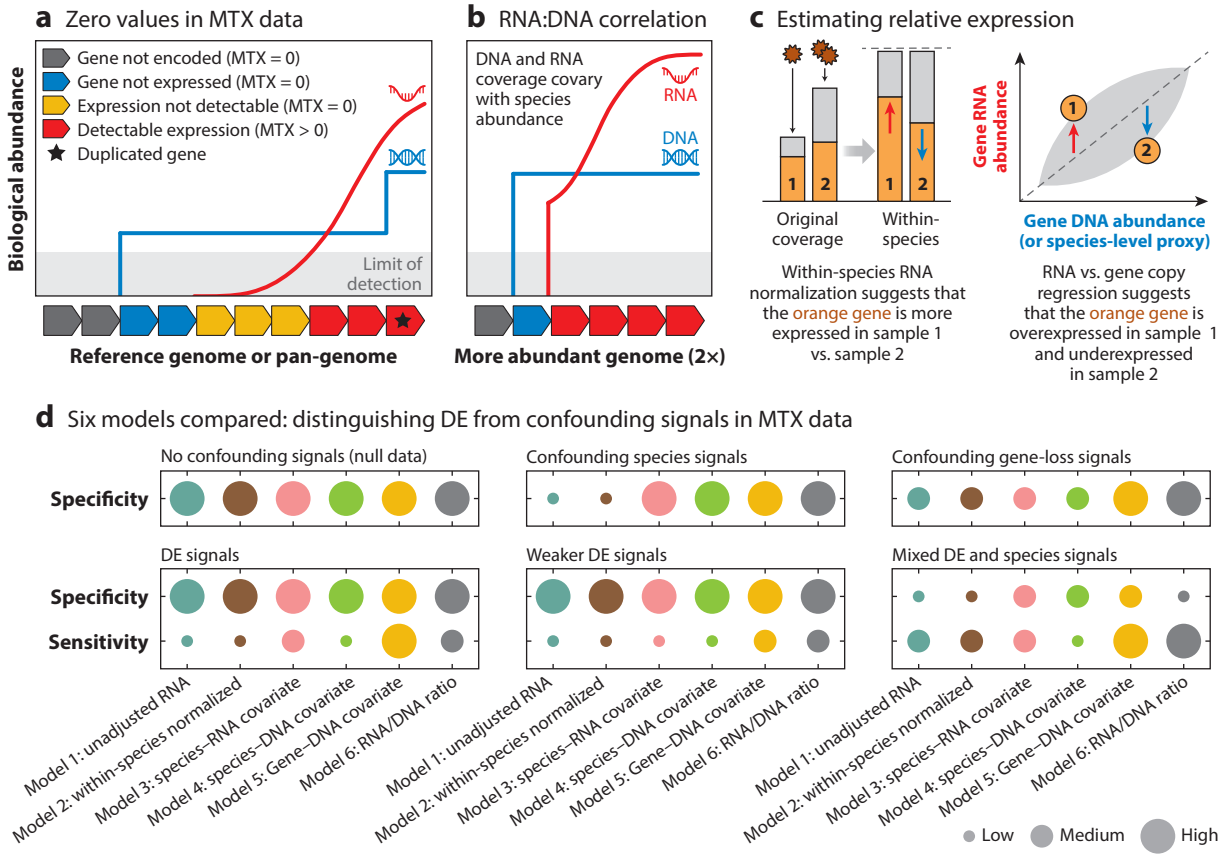

Figure 2

Statistical considerations for meta-omic DE analysis. (*a*) Like single-organism RNA-seq, MTX zeroes can represent biological nonexpression or technical effects (e.g., expression below the limit of detection of a given sequencing depth). MTX zeroes can also result from strain-specific gene loss. (*b*) As a species increases in MGX abundance, per-gene MTX abundance increases in kind, but does not indicate a change in gene expression pattern (although MTX zeroes due to nondetection may become less prevalent). (*c*) Identifying true changes in community functional activity requires adjusting for DNA gene copy number (i.e., relative expression). Possible approaches include within-species RNA normalization and per-gene normalization against an estimate of gene copy number. (*d*) An evaluation of six models (*x*-axes) for detecting DE from a microbial community context. Each panel corresponds to a synthetic human gut dataset and a random case/control phenotype permuted to associate with nothing (null data), species abundance, gene loss, or gene expression. Specificity is higher (*larger circles*) when a model reports false positives at the expected rate for a given dataset (nominal p value < 0.05). Sensitivity (*bottom panels only*) is higher when a model reports a larger fraction of spiked gene expression signals from a given dataset (FDR-adjusted p value < 0.05). Models adjusting RNA for DNA copy number as either a ratio (6) or, especially, covariate (5) were generally most sensitive and specific. Abbreviations: DE, differential expression; FDR, false discovery rate; RNA-seq, RNA sequencing.

by its DNA abundance. This procedure is feasible for genes of unknown taxonomy, and it further accounts for gene loss and duplication events. A hybrid approach involves performing this normalization using a proxy for gene abundance, such as the taxonomic abundance of the gene's encoding species or (in the absence of MGX data) an estimate of taxonomic abundance derived from total RNA contributions. Like within-species RNA normalization, these approaches require knowledge of gene–species relationships and assume a constant gene copy number, but may be more numerically stable when species abundance is low.

Differences in the effectiveness of these approaches for identifying differential gene expression in microbial communities are apparent from simulated meta-omic sequencing data. As an example, the healthy human gut microbiome (16) can be used as a community template for synthetic functional profiles by sampling transcripts within community species following a log-normal abundance distribution. Datasets containing 100 such profiles were then permuted to induce correlations between gene expression and a randomly generated case–control phenotype (thus serving as gold standard differential expression relationships to recover). These synthetic correlations were complemented with additional datasets containing (*a*) no phenotypic associations, (*b*) confounded associations between phenotype and species abundance, and (*c*) confounded associations between phenotype and gene loss (which serve as potential sources of spurious expression-level associations). Before performing statistical testing on these datasets, we sum-normalized their synthetic abundances to CPM units (80) to adjust for differences in sequencing depth and log-transformed abundances for variance stabilization. Samples with no DNA or RNA counts for a given gene were treated as uninformative and ignored; remaining zero values were additively smoothed by half the gene's smallest nonzero measurement.

Six different statistical models for differential MTX analysis showed strikingly different behaviors on these synthetic communities (**Figure 2*d***). Each model treated transcript abundance as a dependent variable and phenotype as a fixed covariate of interest. Model 1 made no attempt to account for DNA-relative expression, while five other models optionally included within-species RNA normalization (models 2 and 3), a species-total RNA covariate (model 3), a species-total DNA covariate (model 4), a per-gene DNA covariate (model 5), or a per-gene RNA/DNA ratio (model 6). The naive model (model 1) was particularly prone to false positives when transcripts' source species varied with phenotype. Surprisingly, this was also the case for the within-species normalization model (model 2), possibly due to issues from species with low absolute RNA abundance. Models 1–4 struggled to distinguish gene-loss events from reduced expression, while models incorporating a DNA-level estimate of gene copy number (models 5 and 6) maintained specificity under those circumstances. Incorporating DNA-level gene abundance as a covariate (model 5) proved to be the most effective approach for identifying signatures of differential expression, even when combined with simultaneous changes in species abundance. Using a per-gene RNA/DNA ratio as a measure of relative expression (model 6) was also reasonably sensitive, but more prone to false positives (reduced specificity) when gene expression and species abundance jointly associated with phenotype. Within-species RNA normalization with a separate total species RNA covariate provided the best balance of sensitivity and specificity in the absence of DNA data (although it does require foreknowledge of gene–species relationships).

MICROBIAL COMMUNITY TRANSCRIPTION IN THE HUMAN MICROBIOME

Many applications of MTX to date have been in the human microbiome in order to improve our basic biological understanding of these important ecologies and to better apply them to health and disease (**Figure 3**) (80, 134). Some of the most basic applications of MTX to the human microbiome first demonstrated that, as expected from single organisms, transcriptional patterns change rapidly but, like metagenomes, are also highly individualized (9, 12, 24, 134). For the few microbes detectable across different body areas, transcriptional activity was also observed to be site specific (9). More recent studies have used MTX to identify host–microbe interactions (14), define core metabolic processes across the body (80), differentiate core versus variably transcribed functions (12), or determine the impact of expressed functions on disease progression (69) and

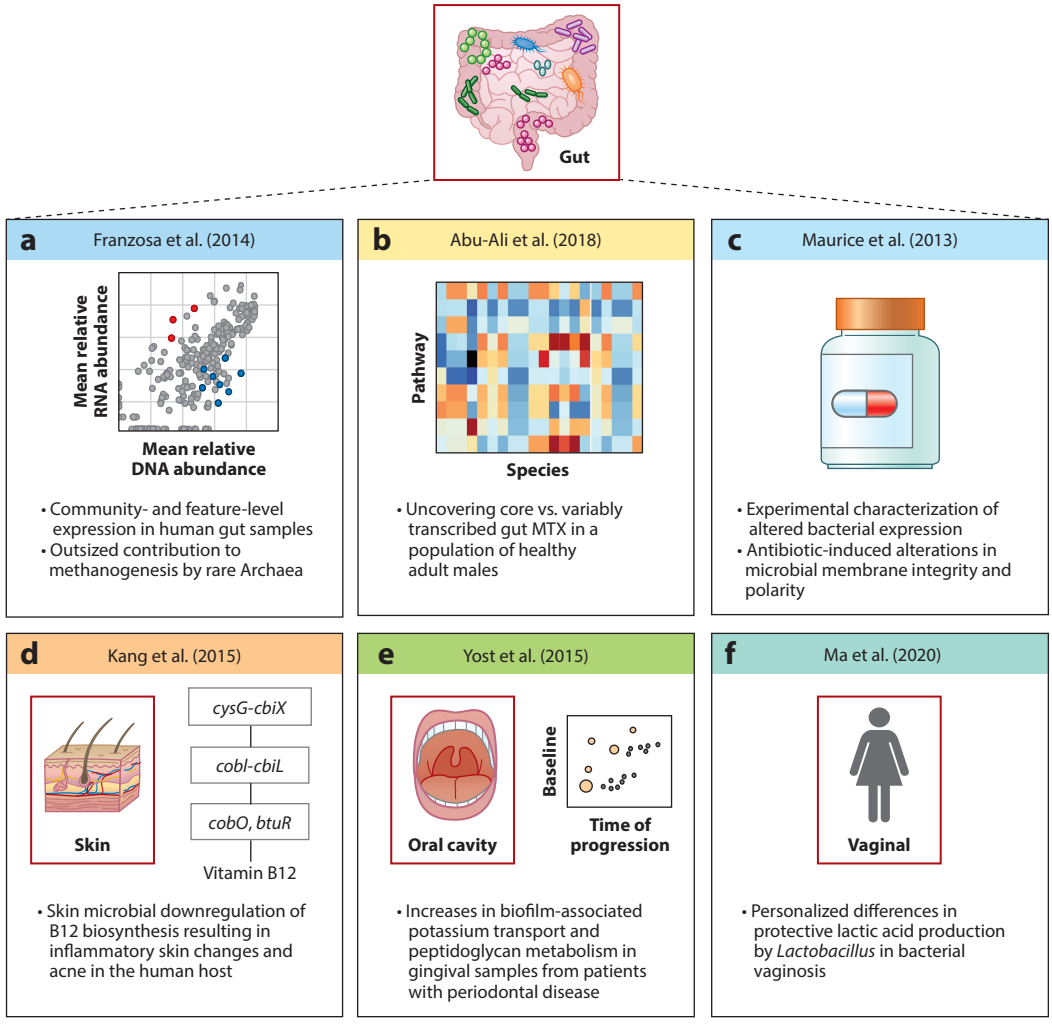

Figure 3

Human microbiome MTX epidemiology. Selected highlights from investigations exploring human-associated MTX profiles demonstrate the potential of future work in the field. Examples include (*a*) a human gut MTX study that identified metabolism unique to specific, low-abundance clades, such as methanogenesis by Archaea (9); (*b*) a study characterizing a core (e.g., amino acid biosynthesis and carbohydrate degradation) versus variable (e.g., sulfate reduction or biotin biosynthesis) gut metatranscriptome in a generally healthy population of older men (12); (*c*) experimental approaches to interrogating how xenobiotics alter not only overall community structure but also the expression of microbial stress response pathways (e.g., the cell death regulator MazG) (146); (*d*) a case–control study of the skin microbiome that found reduced *Propionibacterium acnes* transcription of vitamin B12 synthesis genes—a direct response to exogenous B12 supplementation—resulting in downstream production of proinflammatory, acne-inducing porphyrins (163); (*e*) a sampling of subgingival biofilms that revealed functional signatures of oral dysbiosis linked to periodontal disease progression (153); and (*f*) the development of comprehensive curated gene catalogs to offer greater insights into the complex interplay between downregulation of host defense mechanisms and susceptibility to dysbiosis-associated conditions such as bacterial vaginosis (169).

severity over time (6). Community-derived isolate transcriptomics have a long history, of course (32, 135, 136), and have also recently been used in combination with high-throughput culturomics to functionally characterize, e.g., anaerobic gut commensals (137). In addition, MTX can detect and quantify RNA viruses (138, 139), which are otherwise excluded from MGX surveys.

Like MGX, human microbiome MTX has been used most often for the gut microbiome. This has often focused on inflammatory disorders, with metatranscriptional profiling of conditions spanning inflammatory bowel disease (6, 140), colorectal cancer development (141–143), type 1 diabetes (7), melanoma (144), and other disease phenotypes (145). Additionally, MTX has also been used to investigate biochemical processes in biotransformation of xenobiotics (146) or nutrient utilization (12), e.g., biosynthesis of L-isoleucine and L-tryptophan, polysaccharide (diet and mucous) degradation, and fermentation. These processes directly underlie microbial responses to the most rapid common perturbations of the gut community: response to dietary interventions (147–149) such as fermented milk products (8) or the intake of medications (146), even when only transcription (MTX) and not community structure (MGX) responds to some such stimuli.

Although arguably less well studied than the gut, similar examples exist of in situ microbial gene expression patterns from the oral cavity (9, 150, 151). MTX of the oral microbiome has provided insights into functions and active community members in caries (151, 152), periodontitis (153), biofilm formation, and meal digestion (154). As with the gut, the primary examples examined in community contexts are those known from pathogens previously studied in isolation (155–157), such as proteolysis transcripts, oligopeptide transport, iron metabolism, and flagellum assembly, as contributed by the members of the oral microbiome such as *Fusobacteria* (158, 159). Likewise, longitudinal combinations of MXG and MTX have exposed *Porphyromonas gingivalis* virulence signatures, such as cobalamin biosynthesis, proteolysis, and potassium transport, in initiating progression to periodontitis (153). MTX of periodontitis-associated subgingival biofilms similarly found upregulation of proteolysis, peptide transport, iron acquisition, and lipopolysaccharide synthesis, all of which likely accelerate the proinflammatory potential of the community (156). Some noncoding RNA studies have even been carried out in the oral community context (160), identifying over 40 small RNA families differentially regulated during periodontitis (e.g., amino acid and carbohydrate metabolism, ethanolamine catabolism, stress response).

Due to the biogeographical heterogeneity of the skin, there are considerable challenges in acquiring MTX samples from topographical surfaces that are both targeted and of sufficiently high biomass (161). Although most studies have thus combined MGX with host transcriptional profiling as a result (162), others have obtained sufficient skin microbial material for detection of, e.g., metabolite-mediated responses in acne patients to changes in diet, especially vitamin B12 biosynthesis (163). The transcriptional activities detected in isolates of *Staphylococcus epidermidis* in vitro, such as the inhibition of biofilm formation and antimicrobial peptide expression, may also protect normal skin in a healthy state (164, 165). More detailed transcriptomic (166) or metabolomic (167) profiles will thus likely continue to provide a more complete understanding of the diverse antimicrobial defense systems of the skin.

MTX is also technically challenging in the vaginal microbiome, but for a quite different reason, as samples typically contain relatively low levels of microbial RNA compared to an overabundance of host RNA (168). Among the few studies that have addressed this primary hurdle, transcription in most cases is derived from the community's dominant lactobacilli (169). As expected from earlier work, secretion of antagonistic substances such as lactic acid, H_2O_2, and bacteriocins (170, 171) was common to maintain the clade's community dominance, as were other defense mechanisms during dysbiosis related to bacterial vaginosis (BV), such as cytolysin activation, mucin and glycerol transport, and CRISPR systems (172). In examples from even more severe infectious diseases, the

activated genes responsible for the DNA damage repair in *Gardnerella vaginalis* can contribute not only to BV but also to risk of sexually transmitted infections such as HIV (70).

METATRANSCRIPTOMICS IN NONHUMAN AND FREE-LIVING COMMUNITIES

Research into the transcriptional profiles of nonhuman, host-associated, and, especially, free-living communities (e.g., water, soil, and sediment-borne microbes) began several years before human-associated studies, driven by the ease of sample collection, experimental ability and applications, and the potential for industrial advances. Highly abundant transcripts from nonhuman and environmental samples contain both identifiable and novel functional genes for bioprospecting (30, 50, 173). Therefore, to better leverage microbial functional potential, early studies used molecular techniques developed for MGX to clone DNA from diverse communities into laboratory strains and identify expressed functions through screening, such as color change assays or clearing zones around colonies (174–176). While this approach was able to identify numerous novel enzymes, it proved to be low throughput, and of course it does not scale easily to most community processes (177). This highlights the need for a high-throughput method such as MTX to assess a community's true functional potential and niche-specific differences in transcription.

While in many cases easier to carry out, such nonhuman microbiome research can conversely be challenging to analyze due to the underrepresentation of free-living and host-associated (especially nonmammalian host) isolates in reference databases (30, 173). Free-living communities can also be especially complex, requiring unusually deep sequencing in, e.g., soil and sediment communities (19, 178), and highly novel or diverged organisms are even more common in these environments (30, 50, 173). For instance, Salazar et al. (30) found that only 5% of metagenomes and 10% of MTX reads in the *Tara* Oceans data were taxonomically identifiable using preexisting databases; that number increased to almost 75% of sequences when compared among *Tara* Oceans samples. However, when this self-curated database was used on similar oceanic samples, it was only able to taxonomically identify 43% of the cDNA (49), indicating a large portion of sequences from unknown organisms even with only a small shift in the sampling environment. Several other MTX-based surveys in nonhuman studies have also found low representation of reference sequences (173, 179, 180). New methodologies for assembling or aligning nonhuman and environmental samples often find increases in the number of unknown community members and functions (173), thus indicating that we may not yet know the true depth of the microbial dark matter. Universal microbial housekeeping processes have been detected across all biomes and hosts, including carbohydrate, amino acid, nucleotide, and energy metabolism (173); however, unraveling the preponderance of unique functions in such environments has presented its own set of challenges.

Nonhuman Host-Associated Microbial Communities Display Unique Gene Expression

Although a core set of conserved functions span host-associated microbial communities across the animal kingdom, gene expression has been shown to differ by host, diet, ecosystem, and broader phylogeny (173, 179) (**Tables 1** and **2**). In particular, differences have been observed between mammalian and nonmammal hosts (173, 179, 181) and among omnivores, herbivores, and carnivores (148, 182, 183). Gnotobiotic mouse models were among the first organisms leveraged to study community-wide host-associated expression and to establish MTX protocols (8, 149, 173, 180, 184). With a focus on perturbations relevant to human disease, these studies

Table 1 Mammalian (nonhuman) host-associated MTX studies

Host	References	Major transcribers	Major functions	Major findings
Mice	8, 52, 141, 149, 180, 184–189, 192	Lachnospiracae, Clostridiaceae, Bacteroidaceae, Erysipelotrichaceae, Lactobacillaceae	Carbohydrate, amino acid, and nucleotide metabolism; pyruvate formate-lyase and ABC-type sugar transport systems; butyrate and propionate formation; vitamin B12 biosynthesis; cell motility; regulation; cell signaling; oxidative phosphorylation; lipopolysaccharide production; antibiotic resistance	■ High abundance of carbohydrate metabolism–related transcripts (8, 52, 185, 188, 189) ■ Pyruvate formate-lyase and ABC-type sugar transport systems upregulated in Western diets (149) ■ β-glucosidase and malate-lactate dehydrogenase upregulated in low-fat, plant polysaccharide–rich diets (149) ■ Upregulation of lipopolysaccharide production in response to pollutants and disease (141, 186) ■ Decreased butyrate production, increased oxidative stress response during inflammation and disease states (141, 185)
Nonhuman primates	193, 195, 197	*Prevotella*, *Bacteroides*, *Treponema*, *Ruminococcus*, *Clostridium*, *Megasphaera*, *Eubacterium*, Streptococcaceae, Pasteurellaceae	Carbohydrate, protein, host-derived glycan, and sulfur metabolism; vitamin B synthesis; membrane transport	■ High abundance of carbohydrate metabolism–related transcripts (193, 195) ■ Increased sulfur metabolism, decreased dormancy/sporulation and iron acquisition in inflammatory disease (193)
Pigs	187, 191, 196	*Prevotella*, *Oscillibacter*, other Clostridia, *Bacteriodes*	Carbohydrate, protein, amino acid, and cell wall metabolism; stress response; respiration; virulence; antibiotic resistance	■ Expression of antibiotic resistance genes including macrolides, tetracyclines, aminoglycosides (187, 191)
Sheep	198, 212	*Sharpea*, *Megasphaera*, Euryarchaeota	Amino acid and methionine biosynthesis, galactose metabolism, sugar transport, pyruvate fermentation, butyrate formation	■ Sheep with lower methane production were enriched in transcripts related to short-chain fatty acid production, sugar transport, and amino acid biosynthesis (198)
Cows	199, 213	Prevotellaceae, Succinivibrionaceae, *Methanobrevibacter*, *Methanosphaera*, Thermoplasmata	Methanogenesis; carbohydrate, amino acid, and energy metabolism; translation	■ Methanogenesis decreased with dietary supplementation of unsaturated fatty acids (199)

Although MTX studies to date have been more limited than MGX, particularly outside of the human microbiome, they have spanned other mammalian host-associated communities. These include other primates, mice, and livestock. Core housekeeping functions (e.g., carbohydrate and amino acid metabolism) tend to reflect broad community similarities across most hosts.

revealed patterns of transcriptional regulation associated with, e.g., oxidative stress during colitis (185), as well as metabolic responses to niche perturbations (e.g., dietary changes or pollutants) (52, 149, 186, 187). In healthy mice, MTX studies have shown that genes involved in carbohydrate metabolism and cellular maintenance (e.g., translation) are highly expressed, while stress response genes are less expressed compared to diseased groups—hallmarks of a generally stable, replicating microbiome (52, 185, 188). In mice bred to model inflammatory disease, there have been notable reductions in the community transcription of the genes involved in the regulation of butyrate synthesis and of flagellin-encoding genes (potentially selecting against beneficial commensals), as well as upregulation of mucin degradation pathways (52, 141). A key takeaway of these studies was often that MTX revealed subtler shifts in community behavior than those visible in structural

Table 2 Nonmammalian host-associated MTX studies

Host	References	Major transcribers	Major functions	Major findings
Birds	181, 191, 211	Firmicutes, Proteobacteria	Antibiotic resistance	■ Microbial resistance in all bird orders sampled from both remote and urban locations (181) ■ MTX as surveillance tool for identification of blood-borne parasites in bird vectors, identified pathogens in complex infections and at low intensities (211)
Fish	53, 203, 204	Fusobacteria, Firmicutes, Proteobacteria, Bacteroidetes, Crenarchaeota, Aquificae, Chlorobi, Thermotogae, *Epulopiscium*	Carbohydrate, amino acid, nucleotide, and lipid metabolism and transport; translation; ribosomal structure and biogenesis; energy production and conversion; plant polysaccharide digestion	■ Plant polysaccharide digestion highly represented in transcripts (53, 203) ■ Transcriptional regulation by microbial communities specific to host plant food source (53, 203) ■ Differential expression by feeding state (204)
Termites	179, 209	Firmicutes, Spirochaetes, Fibrobacteres	Nitrogen fixation, amino acid biosynthesis, lignocellulose digestion, antifungal activity, nutrient and metabolite transport, carbohydrate metabolism	■ Functional divergence in communities of termites feeding on different food sources (cow dung versus wood) (209)
Mosquitos	210, 214	Subset of detected taxa: Mesoniviridae, Tombusviridae, Reoviridae, Bunyaviridae, Mononegavirales	No molecular functions highlighted (studies focused on viral taxonomy)	■ Host phylogeny more likely to predict viral community diversity than geographic location (214) ■ MTX not as sensitive as other detection methods, but identifies more viruses (210)

MTX analyses have been expanded to communities associated with diverse nonmammalian hosts including birds, fish, and termites. As with transcriptional regulation in individual niche-derived microbial isolates, most short-term perturbations have been found to result in host- or condition-specific gene expression changes that are highly specific to these diverse environments (e.g., plant-specific hydrolases in fish hosts).

remodeling (i.e., MGX), even to perturbations previously thought to be large (e.g., diet or probiotics) (8, 149, 189).

Studies of community transcriptional patterns in other mammalian hosts have aimed to understand gene expression in a variety of basic and disease-model contexts: antibiotic resistance, agricultural pathogens, and adaptation to environmental factors (e.g., seasonal changes, climate change, and toxicant exposures). Antibiotic resistance gene expression has been quantified in many host-associated communities such as pigs, mice, and humans (also in birds, phytoplankton, and soil); the types of antibiotics to which each host confers and the expression levels are dependent on anthropogenic exposure, age, and species (181, 187, 190, 191). In one study, adult mammals displayed a higher diversity of resistance mechanisms compared to human infants, nonmammals, and environmental samples, although relative expression levels of antibiotic resistance genes did not vary among human adult, human infant, mouse, or pig samples (187). In another study, resistance gene diversity varied across avian groups sampled from a range of locations such as Antarctica and Australia, with the highest gene diversity arising from avian groups dwelling in wastewater treatment facilities, suggesting an anthropogenic influence on community acquisition of antibiotic resistance (181). Intriguingly, expression of antibiotic resistance genes was not dependent on the presence of current antibiotic exposure in some studies, but it was shown to induce expression in mouse models (187, 192). Additionally, MTX analysis in nonhuman primates showed that genes associated with mucin degradation were upregulated in a chronic inflammation group compared to healthy controls, consistent with mouse and human models (193, 194). Other notable MTX studies in nonhuman mammalian hosts showed microbial adaptation to oxidative stress when an

exogenous strain was introduced to the ileum of nonhuman primates (195), as well as differential gene expression in the microbiota of mother-fed versus formula-fed young in pigs (196). Functional profiling of communities in ruminant animals is also one of several microbial settings that has provided insights into microbially facilitated greenhouse gas (methane) emissions, an anthropogenic contributor to global warming via agriculture (197–199).

Nonmammalian host-associated communities can be more challenging to translate directly to human health, but instead shed light on bioengineering applications or basic biology. MTX of phylogenetically diverse hosts spanning aquatic animals and insects inform the discovery of mechanisms underlying gene regulation in the contexts of diet, immunity, industrially relevant pathogens, and microbe–microbe interactions (53, 179, 189). Transcription in relation to diet has been quantified in the gut microbiota of fish, where it has been found that carbohydrate and amino acid metabolism were among the most abundant transcripts, with abundance dependent on fiber intake [as in humans and mammalian hosts (200–202)] and on stage of feeding cycle (53, 203, 204). Two herbivorous fish gut microbiome studies found that the most abundant transcripts encoded plant hydrolyzing enzymes, which arise either from a single taxon in marine surgeonfish or from several genera of freshwater grass carp, reflecting highly specific variation both in host plant consumption and in microbial ecological structure (53, 203).

Insect-associated communities are often studied due to their relationships with insect-borne human pathogens and to plant degradation for biofuel production (205–208). A substantial share of insect gut MTX has focused on termites due to their microbially facilitated cellulose and lignin degrading potential (209); interestingly, termites exposed to a fungal pathogen exhibited increased bacterial symbiont transcription of amidohydrolase, an enzyme with antifungal activity (179). MTX analysis in insects can be used for surveillance of host-associated pathogens as well, such as arboviruses or blood-borne parasites, as demonstrated in mosquito vectors and avian hosts, respectively (210, 211). The identification of arboviruses with MTX was not as sensitive as other more commonly used targeted detection methods when analyzing spiked viral loads, but a variety of viruses from four families were detected in this way, essentially trading depth for breadth (210). Nonhuman host-associated studies thus show the diversity of microbial gene expression across host phylogenies in response to common environmental factors and clade-specific pathogens or inform the impact of microbial community functions on the broader environment (e.g., methane emissions, antibiotic-resistant gene transfer).

Free-Living Microbiomes Exhibit Differential Gene Expression by Habitat

Earth's ecosystems host a staggering biomass of free-living microbial species essential to support life. These communities differ from host-associated communities due to several factors, including significant and frequent changes in the surrounding abiotic environment (e.g., temperature, salinity, pH, and nutrient availability), decreased constraints on immigration and emigration of novel microbial components, more complex and heterogeneous compositions (e.g., richer combinations of archaea, viruses, microbial eukaryotes, and indirect interactions with multiple higher-order plants and animals), highly dynamic microbial interactions, and more variable biomass. These communities are essential for the cycling of organic, inorganic, and metallic compounds globally (30, 178, 215). Conversely, some environmental microbes contribute to processes that increase the toxicity of xenobiotics (216), release climate-active compounds (217), or act as a reservoir for antibiotic resistance (218). Therefore, it is vital to study not only the genomic content of free-living communities but also their members' viabilities and behaviors in response to environmental fluctuations. Additionally, functional redundancy within environmental communities is likely low—in contrast to host-associated communities—as studies have found a high correlation between

taxonomic and MTX functional richness (30, 219, 220). Therefore, functional responses within a community are thought to occur via two mechanisms that are less coupled than in host-associated communities: community or molecular control of gene expression, or shifts in composition via migration or growth of microbes with a new set of complementary functional genes. Any such responses to environmental conditions and community requirements are often grouped into three broad ecosystem types: natural (e.g., soil, sediment, and ocean habitats), extreme (e.g., low-pH environments, hydrothermal vents, and subarctic habitats), and polluted (**Figure 4**).

Without microbial processes occurring in natural ecosystems, organic and inorganic compounds cannot be utilized by higher-order organisms. MTX has been applied to various ecosystems, and as expected, these diverse microbiomes transcribe genes that contribute to the cycling of carbon, phosphorous, nitrogen, and sulfur and the production of oxygen; these are often the most abundantly transcribed functions in each community to support life (30, 49, 50, 221, 222). Environmental microbes also often contain a higher proportion of active eukaryotic and archeal members than host-associated communities (30, 47, 49, 221, 223–225), which generally fulfill unique niches with respect to biogeochemical cycling. For example, eukaryotic and archeal components of soil and ocean habitats contribute greatly to nitrogen cycling (30, 221, 224). Crucially, in such settings, MGX taxonomic and functional profiles do not always predict natural community abundance of transcripts. For example, decomposition of organic matter (carbon cycling) is a vital process in many habitats, but in forests, this process was found to be driven by low-abundance species (178).

Environmental communities are often studied in the context of global warming, and studies in hypoxic coastal sediments and the open ocean have found high transcription of genes that can detoxify harmful sulfur molecules and help with the removal of climate-active gases (221, 226). Correspondingly, climate change will likely alter the composition of microbial communities, shifting a community's function by limiting or changing what can grow in a given habitat, as has already been found in some regions (30, 31). In the open ocean, bacterial small noncoding RNAs have been shown to potentially be involved in niche adaptation and gene regulation (25). Thus, it is important to understand not only the natural processes that are occurring but also how and what organisms are contributing those functions.

In extreme ecosystems, microbes have often evolved to survive and flourish in unusual ways, where specialized adaptations enable utilization of molecules other than carbon, water, and oxygen to catalyze energy production (227, 228). Prior MTX investigations of extreme ecosystems include clouds (229), permafrost (217, 230, 231), hydrothermal vents (54, 227, 232–234), and extremely acidic environments (216, 228). Many of these studies found a high abundance of transcripts for stress response, cellular maintenance, and survival strategies such as the production of osmoprotectants and cryoprotectants (229, 230). Communities from the sea floor, hydrothermal vents, and acidic environments are all dominated by phylogenies of bacteria and archaea that are rare in other settings and often chemolithotrophic, e.g., using sulfur, phosphorus, or iron to catalyze the production of ATP (54, 216, 228, 232–234). Both gene plasticity and community composition determine transcriptional profiles in these extreme, highly adapted communities, as rare community taxa are frequently identified as providing the majority of transcripts for essential elemental cycling (216, 227). These extreme environments can again also provide useful insights into mechanisms driving climate change, such as the increased production of methane by bacterial communities in permafrost as they shift from survival to decomposition during thawing (217, 230).

Anthropogenic contributions to microbial environments in the form of pollution have, unfortunately, driven many recent MTX studies (228, 235). These include oil spills (10, 27, 236, 237), acid mine drainage (216, 228), and persistent environmental pollutants such as perfluorinated compounds (238), pesticides (27, 239, 240), and metal pollutants (239, 241), which often shift both the composition of the community and its most active functions (10, 228, 241). Perhaps the largest

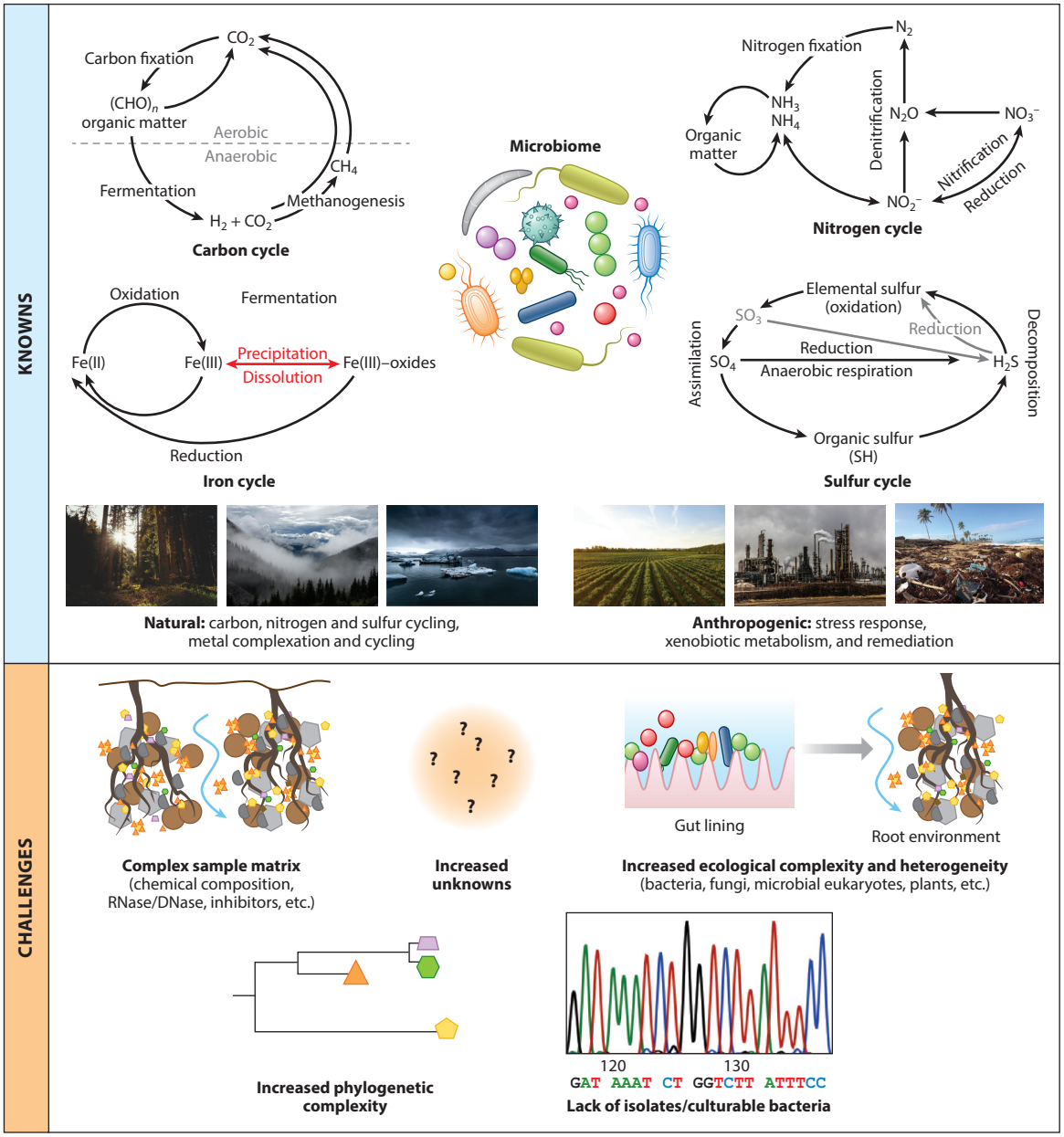

Figure 4

Known roles (*top*) and challenges (*bottom*) in free-living microbial community transcriptional profiling. Free-living microbial communities are essential for the cycling of carbon, nitrogen, sulfur, and metals. These processes are highly transcribed in natural and anthropogenic chemical–associated communities. When in polluted environments, either natively or for remediation, microbes tend to transcribe genes for stress response and xenobiotic metabolism. Environmental communities also present a myriad of challenges to MTX studies, including often having a more diverse composition across multiple kingdoms of life, unusually high or low biomass, increased numbers of phylogenetic and molecular unknowns, a corresponding lack of isolates and reference sequences in databases, and a preponderance of members that are unculturable or very difficult to culture.

pollution event during the age of efficient MTX was the disastrous Deepwater Horizon oil spill in the Gulf of Mexico, during which numerous researchers focused on the potential for microorganisms to remediate oil through natural metabolism to less toxic hydrocarbon derivatives (10, 236). Through these studies, the uncultured order Oceanospirillales was identified as the main transcriber of alkene degradation pathways (10, 236). Interestingly, although microbes encoding proteins with the potential to degrade recalcitrant components of oil were enumerated, they did not appear to be readily transcribing these genes (10).

Similarly, at locations heavily polluted by metals—soil, coastal ecosystems, or abandoned mine sites—the most transcribed functions are often from extremophilic microbes and involved in chemotaxis, energy production, and mitigation strategies against metal toxicity (e.g., ABC transporters or metal-specific membrane-bound exporters) (215, 216, 241, 242). Either energy production or direct mitigation of toxicity can result in detoxification of the pollutant metals and a return to more productive elemental cycling (176, 216, 242). However, microbial biochemical processes can in less beneficial cases also result in the production of toxic chemicals (216). It is thus difficult to predict a priori how metabolically flexible microbial communities will respond when exposed to anthropogenic compounds with which they generally did not evolve—unlike most other chemicals on Earth. This highlights both the massive biochemical potential of free-living microbial communities and the importance of studying these communities using MTX approaches.

WHITHER METATRANSCRIPTOMICS?

With approximately a decade of modern MTX now behind us—and a tremendous range of discoveries to date—it remains to look ahead and ask what challenges persist in leveraging and expanding this still-growing field. Even existing MTX profiles can be difficult to interpret with respect to basic biology, or difficult to apply to biotechnology or human health. Perhaps the greatest hurdle is biological: The half, three-quarters, or in some settings over 90% of highly abundant community transcripts with no biochemical characterization are often unactionable. This could be addressed by a combination of computational techniques [e.g., automated function prediction (243)], high-throughput isolate and community functional screening, and traditional microbiology (244). MTX for the human microbiome still lacks so-called killer apps in the form of biomarkers or treatment targets only visible transcriptionally, and not from MGX or culture-based assays (6). Such discoveries will likely be dependent both on improved MTX data generation and on better bioinformatic and statistical approaches to such data.

Experimentally, platforms for high-fidelity MTX sequencing are less widely available than those for shotgun or amplicon MGX. High-biomass, minimally biased RNA extraction from ecologically and chemically complex communities is an obvious hurdle, although it has improved in recent years due to advancements in commercial kit quality. Depletion of structural (ribosomal and transfer) RNAs is also critical; after a multiyear hiatus in cost-effective kit availability, Illumina's reintroduction of the modified Ribo-Zero Plus kit may help, but it introduces different chemistry and is not equally applicable to all community settings (due to, e.g., phylogenetic and host variability) (57, 245). It is difficult to suggest only one of these technical areas on which to focus since, practically speaking, they must all be solved simultaneously to generate low-noise MTX profiles. Long- and linked-read technologies have benefitted MGX, but especially due to the difficulty of extracting and maintaining high–molecular weight RNA, their applications in MTX remain limited (82). The same is true of microbial single-cell sequencing, which is quite technically challenging even for DNA in communities with mixed cell wall, extracellular matrix, and nuclease compositions (246).

Analytically, the field could also take substantial strides to prepare for the availability of more and higher-quality MTX profiles. While many tools from MGX and single-organism RNA-seq have been applied to MTX sequencing, almost none have been quantitatively evaluated in this setting or developed exclusively for community MTX. What biases with respect to phylogeny or GC (guanine–cytosine) content might be encountered in MTX read quantification? Are sensitivity versus specificity trade-offs in assembly- versus reference-based analyses similar to those encountered in MGX, or uniquely different in MTX? Should MTX transcripts be directly assembled, assembled via mapping to MGX assemblies, neither, or (once possible) both? Downstream of taxonomic and functional feature quantification, MTX statistics must also be evaluated; the misapplication of molecular epidemiology models from RNA-seq to MGX without modification has often resulted in highly inflated false positives (247). The biological questions addressed by models of "How highly transcribed is this gene in a community?" versus "How highly transcribed is it relative to its DNA copy number?" are also not yet well differentiated, as demonstrated above (**Figure 2**). There is not yet even a clear way to determine simply how many samples should be used or how deeply they should be sequenced for MTX power analyses.

The next steps that matter most will be the discovery of applications that can only be enabled via MTX, and not by the variety of other microbial community study methods now available. This might be in human molecular epidemiology, for example, if particular gut MTX configurations (and not MGX profiles) prove to be predictive of infectious disease onset, chronic disease risk, or diet or drug interactions. Especially given the subtlety of interindividual strain diversity, it might be impossible to determine a patient's microbial response to treatment from a taxonomic profile, but monitoring a metatranscriptome could directly measure, e.g., drug metabolism (248). Biofuel production and bioremediation are likewise highly regulation-dependent processes in free-living communities, in which MTX has been preliminarily combined with synthetic biology to engineer desired, sustainable reactions (10, 26). Just as microarrays opened up entirely new avenues in molecular biology, the same remains true for community MTX: There are literally millions of genes out there whose basic intra- and intermicrobial regulatory circuits remain to be understood.

DISCLOSURE STATEMENT

Curtis Huttenhower is a member of the Scientific Advisory Boards of Seres Therapeutics and Empress Therapeutics.

ACKNOWLEDGMENTS

This work was supported in part by NIH NIDDK (National Institute of Diabetes and Digestive and Kidney Diseases) grant R24DK110499 (to C.H.), Takeda Pharmaceuticals research agreement 4100215484 (to C.H.), the Versus Arthritis Inflammatory Arthritis Research Consortium grant R39789 (to Fiona Powrie), and Cancer Research UK Grand Challenge initiative C10674/A27140 (to Wendy Garrett).

LITERATURE CITED

1. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, et al. 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* 269:496–512
2. Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, et al. 1998. The transcriptional program of sporulation in budding yeast. *Science* 282:699–705
3. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. 2017. Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35:833–44

4. Bashiardes S, Zilberman-Schapira G, Elinav E. 2016. Use of metatranscriptomics in microbiome research. *Bioinform. Biol. Insights* 10:19–25
5. Franzosa EA, Hsu T, Sirota-Madi A, Shafquat A, Abu-Ali G, et al. 2015. Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nat. Rev. Microbiol.* 13:360–72
6. Schirmer M, Franzosa EA, Lloyd-Price J, McIver LJ, Schwager R, et al. 2018. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat. Microbiol.* 3:337–46
7. Heintz-Buschart A, May P, Laczny CC, Lebrun LA, Bellora C, et al. 2016. Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nat. Microbiol.* 2:16180
8. McNulty NP, Yatsunenko T, Hsiao A, Faith JJ, Muegge BD, et al. 2011. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. *Sci. Transl. Med.* 3:106ra06
9. Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, et al. 2014. Relating the metatranscriptome and metagenome of the human gut. *PNAS* 111:E2329–38
10. Mason OU, Hazen TC, Borglin S, Chain PS, Dubinsky EA, et al. 2012. Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME J.* 6:1715–27
11. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, et al. 2011. Metagenomic biomarker discovery and explanation. *Genome Biol.* 12:R60
12. Abu-Ali GS, Mehta RS, Lloyd-Price J, Mallick H, Branck T, et al. 2018. Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat. Microbiol.* 3:356–66
13. Human Microbiome Proj. Consort. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207–14
14. Pérez-Losada M, Castro-Nallar E, Bendall ML, Freishtat RJ, Crandall KA. 2015. Dual transcriptomic profiling of host and microbiota during health and disease in pediatric asthma. *PLOS ONE* 10:e0131819
15. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59–65
16. Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, et al. 2017. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* 550:61–66
17. Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, et al. 2016. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N. Engl. J. Med.* 375:717–29
18. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, et al. 2011. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Investig.* 121:2750–67
19. Gifford SM, Sharma S, Rinta-Kanto JM, Moran MA. 2011. Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *ISME J.* 5:461–72
20. Moran MA, Satinsky B, Gifford SM, Luo H, Rivers A, et al. 2013. Sizing up metatranscriptomics. *ISME J.* 7:237–43
21. Staubach F, Baines JF, Künzel S, Bik EM, Petrov DA. 2013. Host species and environmental effects on bacterial communities associated with *Drosophila* in the laboratory and in the natural environment. *PLOS ONE* 8:e70749
22. Gómez P, Buckling A. 2011. Bacteria-phage antagonistic coevolution in soil. *Science* 332:106–9
23. Palková Z. 2004. Multicellular microorganisms: laboratory versus nature. *EMBO Rep.* 5:470–76
24. Bao G, Wang M, Doak TG, Ye Y. 2015. Strand-specific community RNA-seq reveals prevalent and dynamic antisense transcription in human gut microbiota. *Front. Microbiol.* 6:896
25. Shi Y, Tyson GW, DeLong EF. 2009. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* 459:266–69
26. Simmons CW, Reddy AP, D'Haeseleer P, Khudyakov J, Billis K, et al. 2014. Metatranscriptomic analysis of lignocellulolytic microbial communities involved in high-solids decomposition of rice straw. *Biotechnol. Biofuels* 7:495
27. Singh DP, Prabha R, Gupta VK, Verma MK. 2018. Metatranscriptome analysis deciphers multifunctional genes and enzymes linked with the degradation of aromatic compounds and pesticides in the wheat rhizosphere. *Front. Microbiol.* 9:1331

28. Ishii S, Suzuki S, Tenney A, Nealson KH, Bretschger O. 2018. Comparative metatranscriptomics reveals extracellular electron transfer pathways conferring microbial adaptivity to surface redox potential changes. *ISME J*. 12:2844–63
29. Mann E, Wetzels SU, Wagner M, Zebeli Q, Schmitz-Esser S. 2018. Metatranscriptome sequencing reveals insights into the gene expression and functional potential of rumen wall bacteria. *Front. Microbiol.* 9:43
30. Salazar G, Paoli L, Alberti A, Huerta-Cepas J, Ruscheweyh HJ, et al. 2019. Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. *Cell* 179:1068–83.e21
31. Peng J, Wegner CE, Bei Q, Liu P, Liesack W. 2018. Metatranscriptomics reveals a differential temperature effect on the structural and functional organization of the anaerobic food web in rice field soil. *Microbiome* 6:169
32. Giannoukos G, Ciulla DM, Huang K, Haas BJ, Izard J, et al. 2012. Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol*. 13:R23
33. Mallick H, Ma S, Franzosa EA, Vatanen T, Morgan XC, Huttenhower C. 2017. Experimental design and quantitative analysis of microbial community multiomics. *Genome Biol*. 18:228
34. Joice R, Yasuda K, Shafquat A, Morgan XC, Huttenhower C. 2014. Determining microbial products and identifying molecular targets in the human microbiome. *Cell Metab*. 20:731–41
35. Zhou J. 2003. Microarrays for bacterial detection and microbial community analysis. *Curr. Opin. Microbiol.* 6:288–94
36. Mäder U, Nicolas P, Richard H, Bessières P, Aymerich S. 2011. Comprehensive identification and quantification of microbial transcriptomes by genome-wide unbiased methods. *Curr. Opin. Biotechnol.* 22:32–41
37. Siezen RJ, Wilson G, Todt T. 2010. Prokaryotic whole-transcriptome analysis: deep sequencing and tiling arrays. *Microb. Biotechnol.* 3:125–30
38. Sorek R, Cossart P. 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat. Rev. Genet.* 11:9–16
39. Abdelrahman YM, Rose LA, Belland RJ. 2011. Developmental expression of non-coding RNAs in *Chlamydia trachomatis* during normal and persistent growth. *Nucleic Acids Res*. 39:1843–54
40. Kumar R, Shah P, Swiatlo E, Burgess SC, Lawrence ML, Nanduri B. 2010. Identification of novel noncoding small RNAs from *Streptococcus pneumoniae* TIGR4 using high-resolution genome tiling arrays. *BMC Genom*. 11:350
41. Thomassen GO, Weel-Sneve R, Rowe AD, Booth JA, Lindvall JM, et al. 2010. Tiling array analysis of UV treated *Escherichia coli* predicts novel differentially expressed small peptides. *PLOS ONE* 5:e15356
42. Yergeau E, Lawrence JR, Waiser MJ, Korber DR, Greer CW. 2010. Metatranscriptomic analysis of the response of river biofilms to pharmaceutical products, using anonymous DNA microarrays. *Appl. Environ. Microbiol.* 76:5432–39
43. Georg J, Voss B, Scholz I, Mitschke J, Wilde A, Hess WR. 2009. Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Mol. Syst. Biol.* 5:305
44. Parro V, Moreno-Paz M, González-Toril E. 2007. Analysis of environmental transcriptomes by DNA microarrays. *Environ. Microbiol.* 9:453–64
45. Poretsky RS, Bano N, Buchan A, LeCleir G, Kleikemper J, et al. 2005. Analysis of microbial gene transcripts in environmental samples. *Appl. Environ. Microbiol.* 71:4121–26
46. Botero LM, D'Imperio S, Burr M, McDermott TR, Young M, Hassett DJ. 2005. Poly(A) polymerase modification and reverse transcriptase PCR amplification of environmental RNA. *Appl. Environ. Microbiol.* 71:1267–75
47. Leininger S, Urich T, Schloter M, Schwark L, Qi J, et al. 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442:806–9
48. Urich T, Lanzén A, Qi J, Huson DH, Schleper C, Schuster SC. 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLOS ONE* 3:e2527
49. Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, et al. 2008. Microbial community gene expression in ocean surface waters. *PNAS* 105:3805–10

50. Gilbert JA, Field D, Huang Y, Edwards R, Li W, et al. 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLOS ONE* 3:e3042
51. Marchetti A, Schruth DM, Durkin CA, Parker MS, Kodner RB, et al. 2012. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *PNAS* 109:E317–25
52. Schwab C, Berry D, Rauch I, Rennisch I, Ramesmayer J, et al. 2014. Longitudinal study of murine microbiota activity and interactions with the host during acute inflammation and recovery. *ISME J.* 8:1101–14
53. Ngugi DK, Miyake S, Cahill M, Vinu M, Hackmann TJ, et al. 2017. Genomic diversification of giant enteric symbionts reflects host dietary lifestyles. *PNAS* 114:E7592–601
54. Lopez-Fernandez M, Broman E, Simone D, Bertilsson S, Dopson M. 2019. Statistical analysis of community RNA transcripts between organic carbon and geogas-fed continental deep biosphere groundwaters. *mBio* 10(4):e01470-19
55. Beitel CW, Froenicke L, Lang JM, Korf IF, Michelmore RW, et al. 2014. Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. *PeerJ* 2:e415
56. Bronzato Badial A, Sherman D, Stone A, Gopakumar A, Wilson V, et al. 2018. Nanopore sequencing as a surveillance tool for plant pathogens in plant and insect tissues. *Plant Dis.* 102:1648–52
57. Petrova OE, Garcia-Alcalde F, Zampaloni C, Sauer K. 2017. Comparative evaluation of rRNA depletion procedures for the improved analysis of bacterial biofilm and mixed pathogen culture transcriptomes. *Sci. Rep.* 7:41114
58. Calgaro M, Romualdi C, Waldron L, Risso D, Vitulo N. 2020. Assessment of statistical methods from single cell, bulk RNA-seq, and metagenomics applied to microbiome data. *Genome Biol.* 21:191
59. Deutscher MP. 2006. Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Res.* 34:659–66
60. Shulman LM, Hindiyeh M, Muhsen K, Cohen D, Mendelson E, Sofer D. 2012. Evaluation of four different systems for extraction of RNA from stool suspensions using MS-2 coliphage as an exogenous control for RT-PCR inhibition. *PLOS ONE* 7:e39455
61. Zoetendal EG, Booijink CC, Klaassens ES, Heilig HG, Kleerebezem M, et al. 2006. Isolation of RNA from bacterial samples of the human gastrointestinal tract. *Nat. Protoc.* 1:954–59
62. Reck M, Tomasch J, Deng Z, Jarek M, Husemann P, Wagner-Döbler I. 2015. Stool metatranscriptomics: a technical guideline for mRNA stabilisation and isolation. *BMC Genom.* 16:494
63. Peano C, Pietrelli A, Consolandi C, Rossi E, Petiti L, et al. 2013. An efficient rRNA removal method for RNA sequencing in GC-rich bacteria. *Microb. Inform. Exp.* 3:1
64. He S, Wurtzel O, Singh K, Froula JL, Yilmaz S, et al. 2010. Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nat. Methods* 7:807–12
65. Sultan M, Amstislavskiy V, Risch T, Schuette M, Dökel S, et al. 2014. Influence of RNA extraction methods and library selection schemes on RNA-seq data. *BMC Genom.* 15:675
66. Pang X, Zhou D, Song Y, Pei D, Wang J, et al. 2004. Bacterial mRNA purification by magnetic capture-hybridization method. *Microbiol. Immunol.* 48:91–96
67. Stewart FJ, Ottesen EA, DeLong EF. 2010. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J.* 4:896–907
68. Stark R, Grzelak M, Hadfield J. 2019. RNA sequencing: the teenage years. *Nat. Rev. Genet.* 20:631–56
69. Nowicki EM, Shroff R, Singleton JA, Renaud DE, Wallace D, et al. 2018. Microbiota and metatranscriptome changes accompanying the onset of gingivitis. *mBio* 9(2):e00575-18
70. Deng ZL, Gottschick C, Bhuju S, Masur C, Abels C, Wagner-Döbler I. 2018. Metatranscriptome analysis of the vaginal microbiota reveals potential mechanisms for protection against metronidazole in bacterial vaginosis. *mSphere* 3(3):e00262-18
71. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, et al. 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569:655–62
72. Hao L, Michaelsen TY, Singleton CM, Dottorini G, Kirkegaard RH, et al. 2020. Novel syntrophic bacteria in full-scale anaerobic digesters revealed by genome-centric metatranscriptomics. *ISME J.* 14:906–18
73. Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS. 2018. Long reads: their purpose and place. *Hum. Mol. Genet.* 27:R234–41

74. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol*. 21:30
75. Leimena MM, Ramiro-Garcia J, Davids M, van den Bogert B, Smidt H, et al. 2013. A comprehensive metatranscriptome analysis pipeline and its validation using human small intestine microbiota datasets. *BMC Genom*. 14:530
76. Martinez X, Pozuelo M, Pascal V, Campos D, Gut I, et al. 2016. MetaTrans: an open-source pipeline for metatranscriptomics. *Sci. Rep.* 6:26447
77. Narayanasamy S, Jarosz Y, Muller EE, Heintz-Buschart A, Herold M, et al. 2016. IMP: a pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses. *Genome Biol*. 17:260
78. Ni Y, Li J, Panagiotou G. 2016. COMAN: a web server for comprehensive metatranscriptomics analysis. *BMC Genom*. 17:622
79. Westreich ST, Treiber ML, Mills DA, Korf I, Lemay DG. 2018. SAMSA2: a standalone metatranscriptome analysis pipeline. *BMC Bioinform*. 19:175
80. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, et al. 2018. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* 15:962–68
81. Tamames J, Puente-Sánchez F. 2018. SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Front. Microbiol.* 9:3349
82. Shakya M, Lo CC, Chain PSG. 2019. Advances and challenges in metatranscriptomic analysis. *Front. Genet.* 10:904
83. Niu SY, Yang J, McDermaid A, Zhao J, Kang Y, Ma Q. 2018. Bioinformatics tools for quantitative and functional metagenome and metatranscriptome data analysis in microbes. *Brief. Bioinform.* 19:1415–29
84. Kim J, Kim MS, Koh AY, Xie Y, Zhan X. 2016. FMAP: functional mapping and analysis pipeline for metagenomics and metatranscriptomics studies. *BMC Bioinform.* 17:420
85. Leung HC, Yiu SM, Parkinson J, Chin FY. 2013. IDBA-MT: de novo assembler for metatranscriptomic data generated from next-generation sequencing technology. *J. Comput. Biol.* 20:540–50
86. Leung HC, Yiu SM, Chin FY. 2015. IDBA-MTP: a hybrid metatranscriptomic assembler based on protein information. *J. Comput. Biol.* 22:367–76
87. Ye Y, Tang H. 2016. Utilizing de Bruijn graph of metagenome assembly for metatranscriptome analysis. *Bioinformatics* 32:1001–8
88. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15:550
89. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–40
90. Altelaar AF, Munoz J, Heck AJ. 2013. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* 14:35–48
91. Lo CC, Chain PS. 2014. Rapid evaluation and quality control of next generation sequencing data with FaQCs. *BMC Bioinform*. 15:366
92. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–20
93. Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28:3211–17
94. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–69
95. Martin JA, Wang Z. 2011. Next-generation transcriptome assembly. *Nat. Rev. Genet.* 12:671–82
96. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357–59
97. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–60
98. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12:59–60
99. Steinegger M, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* 35:1026–28

100. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 44:D733–45
101. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 44:D279–85
102. Gene Ontol. Consort. 2015. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* 43:D1049–56
103. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 44:D457–62
104. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res*. 43:D261–69
105. UniProt Consort. 2015. UniProt: a hub for protein information. *Nucleic Acids Res*. 43:D204–12
106. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, et al. 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* 12:902–3
107. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*. 15:R46
108. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh HJ, et al. 2019. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat. Commun.* 10:1014
109. Cabral DJ, Penumutchu S, Reinhart EM, Zhang C, Korry BJ, et al. 2019. Microbial metabolism modulates antibiotic susceptibility within the murine gut microbiome. *Cell Metab*. 30:800–23.e7
110. Serrano MG, Parikh HI, Brooks JP, Edwards DJ, Arodz TJ, et al. 2019. Racioethnic diversity in the dynamics of the vaginal microbiome during pregnancy. *Nat. Med.* 25:1001–11
111. Cabral DJ, Wurster JI, Korry BJ, Penumutchu S, Belenky P. 2020. Consumption of a Western-style diet modulates the response of the murine gut microbiome to ciprofloxacin. *mSystems* 5(4):e00317-20
112. Bairoch A. 2000. The ENZYME database in 2000. *Nucleic Acids Res*. 28:304–5
113. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, et al. 2014. Pfam: the protein families database. *Nucleic Acids Res*. 42:D222–30
114. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. 2000. Gene ontology: tool for the unification of biology. *Nat. Genet.* 25:25–29
115. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res*. 42:D199–205
116. Ugarte A, Vicedomini R, Bernardes J, Carbone A. 2018. A multi-source domain annotation pipeline for quantitative metagenomic and metatranscriptomic functional profiling. *Microbiome* 6:149
117. Pachter L. 2011. Models for transcript quantification from RNA-Seq. arXiv:1104.3889 [q-bio.GN]
118. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, et al. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res*. 33:5691–702
119. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 23:1282–88
120. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19:455–77
121. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–76
122. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644–52
123. Peng Y, Leung HC, Yiu SM, Lv MJ, Zhu XG, Chin FY. 2013. IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. *Bioinformatics* 29:i326–34
124. Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. 2019. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience* 8(9):giz100
125. Celaj A, Markle J, Danska J, Parkinson J. 2014. Comparison of assembly algorithms for improving rate of metatranscriptomic functional annotation. *Microbiome* 2:39
126. Schulz MH, Zerbino DR, Vingron M, Birney E. 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28:1086–92
127. Robertson G, Schein J, Chiu R, Corbett R, Field M, et al. 2010. De novo assembly and analysis of RNA-seq data. *Nat. Methods* 7:909–12

128. Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, et al. 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief. Bioinform.* 14:671–83
129. Lin Y, Golovnina K, Chen ZX, Lee HN, Negron YL, et al. 2016. Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual *Drosophila melanogaster*. *BMC Genom*. 17:28
130. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31:46–53
131. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, et al. 2015. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 43:e47
132. Tarazona S, Furió-Tarí P, Turrà D, Pietro AD, Nueda MJ, et al. 2015. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res*. 43:e140
133. Klingenberg H, Meinicke P. 2017. How to normalize metatranscriptomic count data for differential expression analysis. *PeerJ* 5:e3859
134. Turnbaugh PJ, Quince C, Faith JJ, McHardy AC, Yatsunenko T, et al. 2010. Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *PNAS* 107:7503–8
135. Nouaille S, Even S, Charlier C, Le Loir Y, Cocaign-Bousquet M, Loubière P. 2009. Transcriptomic response of *Lactococcus lactis* in mixed culture with *Staphylococcus aureus*. *Appl. Environ. Microbiol.* 75:4473–82
136. Rey FE, Faith JJ, Bain J, Muehlbauer MJ, Stevens RD, et al. 2010. Dissecting the in vivo metabolic potential of two human gut acetogens. *J. Biol. Chem.* 285:22082–90
137. Ryan D, Jenniches L, Reichardt S, Barquist L, Westermann AJ. 2020. A high-resolution transcriptome map identifies small RNA regulation of metabolism in the gut microbe *Bacteroides thetaiotaomicron*. *Nat. Commun.* 11:3557
138. Zhang T, Breitbart M, Lee WH, Run JQ, Wei CL, et al. 2006. RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLOS Biol*. 4:e3
139. Culley AI, Lang AS, Suttle CA. 2006. Metagenomic analysis of coastal RNA virus communities. *Science* 312:1795–98
140. Zhang X, Deeke SA, Ning Z, Starr AE, Butcher J, et al. 2018. Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric inflammatory bowel disease. *Nat. Commun.* 9:2873
141. Daniel SG, Ball CL, Besselsen DG, Doetschman T, Hurwitz BL. 2017. Functional changes in the gut microbiome contribute to transforming growth factor β-deficient colon cancer. *mSystems* 2:(5)e00065-17
142. Dutilh BE, Backus L, van Hijum SA, Tjalsma H. 2013. Screening metatranscriptomes for toxin genes as functional drivers of human colorectal cancer. *Best Pract. Res. Clin. Gastroenterol.* 27:85–99
143. Yan Y, Drew DA, Markowitz A, Lloyd-Price J, Abu-Ali G, et al. 2020. Structure of the mucosal and stool microbiome in Lynch syndrome. *Cell Host Microbe* 27:585–600.e4
144. Peters BA, Wilson M, Moran U, Pavlick A, Izsak A, et al. 2019. Relating the gut metagenome and metatranscriptome to immunotherapy responses in melanoma patients. *Genome Med*. 11:61
145. Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, et al. 2020. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature* 579:567–74
146. Maurice CF, Haiser HJ, Turnbaugh PJ. 2013. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell* 152:39–50
147. Nguyen LH, Ma W, Wang DD, Cao Y, Mallick H, et al. 2020. Association between sulfur-metabolizing bacterial communities in stool and risk of distal colorectal cancer in men. *Gastroenterology* 158:1313–25
148. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, et al. 2014. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505:559–63
149. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. 2009. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* 1:6ra14
150. Solbiati J, Frias-Lopez J. 2018. Metatranscriptome of the oral microbiome in health and disease. *J. Dent. Res.* 97:492–500

151. Peterson SN, Meissner T, Su AI, Snesrud E, Ong AC, et al. 2014. Functional expression of dental plaque microbiota. *Front. Cell. Infect. Microbiol.* 4:108
152. Simón-Soro A, Guillen-Navarro M, Mira A. 2014. Metatranscriptomics reveals overall active bacterial composition in caries lesions. *J. Oral Microbiol.* 6:25443
153. Yost S, Duran-Pinedo AE, Teles R, Krishnan K, Frias-Lopez J. 2015. Functional signatures of oral dysbiosis during periodontitis progression revealed by microbial metatranscriptome analysis. *Genome Med.* 7:27
154. Benítez-Páez A, Belda-Ferre P, Simón-Soro A, Mira A. 2014. Microbiota diversity and gene expression dynamics in human oral biofilms. *BMC Genom.* 15:311
155. Lamont RJ, Koo H, Hajishengallis G. 2018. The oral microbiota: dynamic communities and host interactions. *Nat. Rev. Microbiol.* 16:745–59
156. Duran-Pinedo AE, Chen T, Teles R, Starr JR, Wang X, et al. 2014. Community-wide transcriptome of the oral microbiome in subjects with and without periodontitis. *ISME J.* 8:1659–72
157. Liu Y, Shetty AC, Schwartz JA, Bradford LL, Xu W, et al. 2015. New signaling pathways govern the host response to *C. albicans* infection in various niches. *Genome Res.* 25:679–89
158. Yost S, Stashenko P, Choi Y, Kukuruzinska M, Genco CA, et al. 2018. Increased virulence of the oral microbiome in oral squamous cell carcinoma revealed by metatranscriptome analyses. *Int. J. Oral Sci.* 10:32
159. Ahn SH, Chun S, Park C, Lee JH, Lee SW, Lee TH. 2017. Transcriptome profiling analysis of senescent gingival fibroblasts in response to *Fusobacterium nucleatum* infection. *PLOS ONE* 12:e0188755
160. Duran-Pinedo AE, Yost S, Frias-Lopez J. 2015. Small RNA transcriptome of the oral microbiome during periodontitis progression. *Appl. Environ. Microbiol.* 81:6688–99
161. Oh J, Byrd AL, Deming C, Conlan S, Kong HH, Segre JA. 2014. Biogeography and individuality shape function in the human skin metagenome. *Nature* 514:59–64
162. Naik S, Bouladoux N, Wilhelm C, Molloy MJ, Salcedo R, et al. 2012. Compartmentalized control of skin immunity by resident commensals. *Science* 337:1115–19
163. Kang D, Shi B, Erfe MC, Craft N, Li H. 2015. Vitamin B_{12} modulates the transcriptome of the skin microbiota in acne pathogenesis. *Sci. Transl. Med.* 7:293ra103
164. Iwase T, Uehara Y, Shinji H, Tajima A, Seo H, et al. 2010. *Staphylococcus epidermidis* Esp inhibits *Staphylococcus aureus* biofilm formation and nasal colonization. *Nature* 465:346–49
165. Lai Y, Di Nardo A, Nakatsuji T, Leichtle A, Yang Y, et al. 2009. Commensal bacteria regulate Toll-like receptor 3-dependent inflammation after skin injury. *Nat. Med.* 15:1377–82
166. Choi M, Seo IC, Kim DH, Choi J-W, Park JM, Oh JW. 2019. Microbiome profile of psoriasis and atopic dermatitis using transcriptome analysis. *J. Allergy Clin. Immunol.* 143(2):AB434
167. Kaddurah-Daouk R, Baillie RA, Zhu H, Zeng ZB, Wiest MM, et al. 2011. Enteric microbiome metabolites correlate with response to simvastatin treatment. *PLOS ONE* 6:e25482
168. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. 2012. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLOS ONE* 7:e33865
169. Ma B, France MT, Crabtree J, Holm JB, Humphrys MS, et al. 2020. A comprehensive non-redundant gene catalog reveals extensive within-community intraspecies diversity in the human vagina. *Nat. Commun.* 11:940
170. Tachedjian G, O'Hanlon DE, Ravel J. 2018. The implausible "in vivo" role of hydrogen peroxide as an antimicrobial factor produced by vaginal microbiota. *Microbiome* 6:29
171. O'Connell CM, Brochu H, Girardi J, Harrell E, Jones A, et al. 2019. Simultaneous profiling of sexually transmitted bacterial pathogens, microbiome, and concordant host response in cervical samples using whole transcriptome sequencing analysis. *Microb. Cell* 6:177–83
172. Macklaim JM, Fernandes AD, Di Bella JM, Hammond JA, Reid G, Gloor GB. 2013. Comparative meta-RNA-seq of the vaginal microbiota and differential expression by *Lactobacillus iners* in health and dysbiosis. *Microbiome* 1:12
173. Jiang Y, Xiong X, Danska J, Parkinson J. 2016. Metatranscriptomic analysis of diverse microbial communities reveals core metabolic pathways and microbiome-specific functionality. *Microbiome* 4:2
174. Ferrer M, Beloqui A, Timmis KN, Golyshin PN. 2009. Metagenomics for mining new genetic resources of microbial communities. *J. Mol. Microbiol. Biotechnol.* 16:109–23

175. Rashamuse K, Magomani V, Ronneburg T, Brady D. 2009. A novel family VIII carboxylesterase derived from a leachate metagenome library exhibits promiscuous β-lactamase activity on nitrocefin. *Appl. Microbiol. Biotechnol.* 83:491–500
176. Lehembre F, Doillon D, David E, Perrotto S, Baude J, et al. 2013. Soil metatranscriptomics for mining eukaryotic heavy metal resistance genes. *Environ. Microbiol.* 15:2829–40
177. Warnecke F, Hess M. 2009. A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. *J. Biotechnol.* 142:91–95
178. Baldrian P, Kolařík M, Stursová M, Kopecký J, Valášková V, et al. 2012. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J.* 6:248–58
179. Peterson BF, Scharf ME. 2016. Metatranscriptome analysis reveals bacterial symbiont contributions to lower termite physiology and potential immune functions. *BMC Genom.* 17:772
180. Xiong X, Frank DN, Robertson CE, Hung SS, Markle J, et al. 2012. Generation and analysis of a mouse intestinal metatranscriptome through Illumina based RNA-sequencing. *PLOS ONE* 7:e36009
181. Marcelino VR, Wille M, Hurt AC, González-Acuña D, Klaassen M, et al. 2019. Meta-transcriptomics reveals a diverse antibiotic resistance gene pool in avian microbiomes. *BMC Biol.* 17:31
182. Milani C, Alessandri G, Mancabelli L, Mangifesta M, Lugli GA, et al. 2020. Multi-omics approaches to decipher the impact of diet and host physiology on the mammalian gut microbiome. *Appl. Environ. Microbiol.* 86(23):e01864-20
183. Wang F, Wan Y, Yin K, Wei Y, Wang B, et al. 2019. Lower circulating branched-chain amino acid concentrations among vegetarians are associated with changes in gut microbial composition and function. *Mol. Nutr. Food Res.* 63:e1900612
184. Davids M, Hugenholtz F, Martins dos Santos V, Smidt H, Kleerebezem M, Schaap PJ. 2016. Functional profiling of unfamiliar microbial communities using a validated de novo assembly metatranscriptome pipeline. *PLOS ONE* 11:e0146423
185. Ilott NE, Bollrath J, Danne C, Schiering C, Shale M, et al. 2016. Defining the microbial transcriptional response to colitis through integrated host and microbiome profiling. *ISME J.* 10:2389–404
186. Nichols RG, Zhang J, Cai J, Murray IA, Koo I, et al. 2019. Metatranscriptomic analysis of the mouse gut microbiome response to the persistent organic pollutant 2,3,7,8-tetrachlorodibenzofuran. *Metabolites* 10(1):1
187. Versluis D, D'Andrea MM, Ramiro Garcia J, Leimena MM, Hugenholtz F, et al. 2015. Mining microbial metatranscriptomes for expression of antibiotic resistance genes under natural conditions. *Sci. Rep.* 5:11981
188. Chung YW, Gwak HJ, Moon S, Rho M, Ryu JH. 2020. Functional dynamics of bacterial species in the mouse gut microbiome revealed by metagenomic and metatranscriptomic analyses. *PLOS ONE* 15:e0227886
189. Hugenholtz F, Davids M, Schwarz J, Müller M, Tomé D, et al. 2018. Metatranscriptome analysis of the microbial fermentation of dietary milk proteins in the murine gut. *PLOS ONE* 13:e0194066
190. Bikel S, Valdez-Lara A, Cornejo-Granados F, Rico K, Canizales-Quinteros S, et al. 2015. Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. *Comput. Struct. Biotechnol. J.* 13:390–401
191. Wang Y, Hu Y, Liu F, Cao J, Lv N, et al. 2020. Integrated metagenomic and metatranscriptomic profiling reveals differentially expressed resistomes in human, chicken, and pig gut microbiomes. *Environ. Int.* 138:105649
192. Korry BJ, Cabral DJ, Belenky P. 2020. Metatranscriptomics reveals antibiotic-induced resistance gene expression in the murine gut microbiota. *Front. Microbiol.* 11:322
193. Westreich ST, Ardeshir A, Alkan Z, Kable ME, Korf I, Lemay DG. 2019. Fecal metatranscriptomics of macaques with idiopathic chronic diarrhea reveals altered mucin degradation and fucose utilization. *Microbiome* 7:41
194. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, et al. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* 13:R79

195. Golomb BL, Hirao LA, Dandekar S, Marco ML. 2016. Gene expression of *Lactobacillus plantarum* and the commensal microbiota in the ileum of healthy and early SIV-infected rhesus macaques. *Sci. Rep.* 6:24723
196. Poroyko V, White JR, Wang M, Donovan S, Alverdy J, et al. 2010. Gut microbial gene expression in mother-fed and formula-fed piglets. *PLOS ONE* 5:e12459
197. Shi W, Moon CD, Leahy SC, Kang D, Froula J, et al. 2014. Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res.* 24:1517–25
198. Kamke J, Kittelmann S, Soni P, Li Y, Tavendale M, et al. 2016. Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a *Sharpea*-enriched microbiome characterised by lactic acid formation and utilisation. *Microbiome* 4:56
199. Poulsen M, Schwab C, Jensen BB, Engberg RM, Spang A, et al. 2013. Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nat. Commun.* 4:1428
200. Morrison KE, Jašarević E, Howard CD, Bale TL. 2020. It's the fiber, not the fat: significant effects of dietary challenge on the gut microbiome. *Microbiome* 8:15
201. Haenen D, Zhang J, Souza da Silva C, Bosch G, van der Meer IM, et al. 2013. A diet high in resistant starch modulates microbiota composition, SCFA concentrations, and gene expression in pig intestine. *J. Nutr.* 143:274–83
202. Makki K, Deehan EC, Walter J, Bäckhed F. 2018. The impact of dietary fiber on gut microbiota in host health and disease. *Cell Host Microbe* 23:705–15
203. Wu S, Ren Y, Peng C, Hao Y, Xiong F, et al. 2015. Metatranscriptomic discovery of plant biomass-degrading capacity from grass carp intestinal microbiomes. *FEMS Microbiol. Ecol.* 91:fiv107
204. Parris DJ, Morgan MM, Stewart FJ. 2019. Feeding rapidly alters microbiome composition and gene transcription in the clownfish gut. *Appl. Environ. Microbiol.* 85(3):e02479-18
205. Dennison NJ, Jupatanakul N, Dimopoulos G. 2014. The mosquito microbiota influences vector competence for human pathogens. *Curr. Opin. Insect. Sci.* 3:6–13
206. Hughes GL, Dodson BL, Johnson RM, Murdock CC, Tsujimoto H, et al. 2014. Native microbiome impedes vertical transmission of *Wolbachia* in *Anopheles* mosquitoes. *PNAS* 111:12498–503
207. Shi W, Xie S, Chen X, Sun S, Zhou X, et al. 2013. Comparative genomic analysis of the microbiome of herbivorous insects reveals eco-environmental adaptations: biotechnology applications. *PLOS Genet.* 9:e1003131
208. Jing TZ, Qi FH, Wang ZY. 2020. Most dominant roles of insect gut bacteria: digestion, detoxification, or essential nutrient provision? *Microbiome* 8:38
209. He S, Ivanova N, Kirton E, Allgaier M, Bergin C, et al. 2013. Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLOS ONE* 8:e61126
210. Batovska J, Mee PT, Lynch SE, Sawbridge TI, Rodoni BC. 2019. Sensitivity and specificity of metatranscriptomics as an arbovirus surveillance tool. *Sci. Rep.* 9:19398
211. Galen SC, Borner J, Williamson JL, Witt CC, Perkins SL. 2020. Metatranscriptomics yields new genomic resources and sensitive detection of infections for diverse blood parasites. *Mol. Ecol. Resour.* 20:14–28
212. He B, Jin S, Cao J, Mi L, Wang J. 2019. Metatranscriptomics of the Hu sheep rumen microbiome reveals novel cellulases. *Biotechnol. Biofuels* 12:153
213. Li F, Guan LL. 2017. Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle. *Appl. Environ. Microbiol.* 83(9):e00061-17
214. Shi M, Neville P, Nicholson J, Eden JS, Imrie A, Holmes EC. 2017. High-resolution metatranscriptomics reveals the ecological dynamics of mosquito-associated RNA viruses in Western Australia. *J. Virol.* 91(17):e00680-17
215. Pei Y, Tao C, Ling Z, Yu Z, Ji J, et al. 2020. Exploring novel Cr(VI) remediation genes for Cr(VI)-contaminated industrial wastewater treatment by comparative metatranscriptomics and metagenomics. *Sci. Total Environ.* 742:140435
216. Hua ZS, Han YJ, Chen LX, Liu J, Hu M, et al. 2015. Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. *ISME J.* 9:1280–94

217. Singleton CM, McCalley CK, Woodcroft BJ, Boyd JA, Evans PN, et al. 2018. Methanotrophy across a natural permafrost thaw environment. *ISME J.* 12:2544–58
218. Liu Z, Klümper U, Liu Y, Yang Y, Wei Q, et al. 2019. Metagenomic and metatranscriptomic analyses reveal activity and hosts of antibiotic resistance genes in activated sludge. *Environ. Int.* 129:208–20
219. Galand PE, Pereira O, Hochart C, Auguet JC, Debroas D. 2018. A strong link between marine microbial community composition and function challenges the idea of functional redundancy. *ISME J.* 12:2470–78
220. Fierer N, Ladau J, Clemente JC, Leff JW, Owens SM, et al. 2013. Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. *Science* 342:621–24
221. Vorobev A, Dupouy M, Carradec Q, Delmont TO, Annamalé A, et al. 2020. Transcriptome reconstruction and functional analysis of eukaryotic marine plankton communities via high-throughput metagenomics and metatranscriptomics. *Genome Res.* 30:647–59
222. Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP, Moran MA. 2009. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environ. Microbiol.* 11:1358–75
223. Voss C, Fiore-Donno AM, Guerreiro MA, Peršoh D, Bonkowski M. 2019. Metatranscriptomics reveals unsuspected protistan diversity in leaf litter across temperate beech forests, with Amoebozoa the dominating lineage. *FEMS Microbiol. Ecol.* 95(10):fiz142
224. Damon C, Lehembre F, Oger-Desfeux C, Luis P, Ranger J, et al. 2012. Metatranscriptomics reveals the diversity of genes expressed by eukaryotes in forest soils. *PLOS ONE* 7:e28967
225. Geisen S, Tveit AT, Clark IM, Richter A, Svenning MM, et al. 2015. Metatranscriptomic census of active protists in soils. *ISME J.* 9:2178–90
226. Broman E, Sjöstedt J, Pinhassi J, Dopson M. 2017. Shifts in coastal sediment oxygenation cause pronounced changes in microbial community composition and associated metabolism. *Microbiome* 5:96
227. Anantharaman K, Breier JA, Sheik CS, Dick GJ. 2013. Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. *PNAS* 110:330–35
228. Chen LX, Hu M, Huang LN, Hua ZS, Kuang JL, et al. 2015. Comparative metagenomic and metatranscriptomic analyses of microbial communities in acid mine drainage. *ISME J.* 9:1579–92
229. Amato P, Besaury L, Joly M, Penaud B, Deguillaume L, Delort AM. 2019. Metatranscriptomic exploration of microbial functioning in clouds. *Sci. Rep.* 9:4383
230. Soldatov VE. 1965. [On indications for subarachnoid therapy of tuberculous meningitis patients]. *Probl. Tuberk.* 43:50–55
231. Hultman J, Waldrop MP, Mackelprang R, David MM, McFarland J, et al. 2015. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* 521:208–12
232. Galambos D, Anderson RE, Reveillaud J, Huber JA. 2019. Genome-resolved metagenomics and metatranscriptomics reveal niche differentiation in functionally redundant microbial communities at deep-sea hydrothermal vents. *Environ. Microbiol.* 21:4395–410
233. Fortunato CS, Huber JA. 2016. Coupled RNA-SIP and metatranscriptomics of active chemolithoautotrophic communities at a deep-sea hydrothermal vent. *ISME J.* 10:1925–38
234. Lesniewski RA, Jain S, Anantharaman K, Schloss PD, Dick GJ. 2012. The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. *ISME J.* 6:2257–68
235. Mapelli F, Scoma A, Michoud G, Aulenta F, Boon N, et al. 2017. Biotechnologies for marine oil spill cleanup: indissoluble ties with microorganisms. *Trends Biotechnol.* 35:860–70
236. Yooseph S, Nealson KH, Rusch DB, McCrow JP, Dupont CL, et al. 2010. Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* 468:60–66
237. Gonzalez E, Pitre FE, Pagé AP, Marleau J, Guidi Nissim W, et al. 2018. Trees, fungi and bacteria: tripartite metatranscriptomics of a root microbiome responding to soil contamination. *Microbiome* 6:53
238. Cerro-Gálvez E, Roscales JL, Jiménez B, Sala MM, Dachs J, Vila-Costa M. 2020. Microbial responses to perfluoroalkyl substances and perfluorooctanesulfonate (PFOS) desulfurization in the Antarctic marine environment. *Water Res.* 171:115434
239. Falk N, Reid T, Skoyles A, Grgicak-Mannion A, Drouillard K, Weisener CG. 2019. Microbial metatranscriptomic investigations across contaminant gradients of the Detroit River. *Sci. Total Environ.* 690:121–31

240. Gao B, Bian X, Chi L, Tu P, Ru H, Lu K. 2017. Organophosphate diazinon altered quorum sensing, cell motility, stress response, and carbohydrate metabolism of gut microbiome. *Toxicol. Sci.* 157:354–64
241. Jacquiod S, Nunes I, Brejnrod A, Hansen MA, Holm PE, et al. 2018. Long-term soil metal exposure impaired temporal variation in microbial metatranscriptomes and enriched active phages. *Microbiome* 6:223
242. Tan S, Cheung S, Ho TY, Liu H. 2019. Metatranscriptomics of the bacterial community in response to atmospheric deposition in the Western North Pacific Ocean. *Mar. Genom.* 45:57–63
243. Zhou N, Jiang Y, Bergquist TR, Lee AJ, Kacsoh BZ, et al. 2019. The CAFA challenge reports improved protein function prediction and new functional annotations for hundreds of genes through experimental screens. *Genome Biol.* 20:244
244. Anton BP, Chang YC, Brown P, Choi HP, Faller LL, et al. 2013. The COMBREX project: design, methodology, and initial results. *PLOS Biol.* 11:e1001638
245. Kraus AJ, Brink BG, Siegel TN. 2019. Efficient and specific oligo-based depletion of rRNA. *Sci. Rep.* 9:12281
246. Woyke T, Doud DFR, Schulz F. 2017. The trajectory of microbial single-cell sequencing. *Nat. Methods* 14:1045–54
247. Hawinkel S, Mattiello F, Bijnens L, Thas O. 2019. A broken promise: microbiome differential abundance methods do not control the false discovery rate. *Brief. Bioinform.* 20:210–21
248. Haiser HJ, Gootenberg DB, Chatman K, Sirasani G, Balskus EP, Turnbaugh PJ. 2013. Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. *Science* 341:295–98