Microbiome analysis – from technical advances to biological relevance

Michelle I. Smith[1], Williams Turpin[1,2], Andrea D. Tyler[1], Mark S. Silverberg[1,2] and Kenneth Croitoru[1,2]*

Addresses: [1]Zane Cohen Centre for Digestive Diseases, Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, 600 University Avenue, Room 437, Toronto, ON, Canada, M5G 1X5; [2]Institute of Medical Science, Department of Medicine, University of Toronto, Toronto, ON, Canada, M5S 1A8

* Corresponding author: Kenneth Croitoru (kcroitoru@mtsinai.on.ca)

The electronic version of this article is the complete one and can be found at: http://f1000.com/prime/reports/b/6/51

Abstract

The development of culture-independent techniques and next-generation sequencing has led to a staggering rise in the number of microbiome studies over the last decade. Although it remains important to identify the taxa of microbes present in a variety of environmental samples, including the gut microbiomes of healthy and diseased individuals, the next stage of microbiome research will need to focus on uncovering the role of the microbiome rather than its mere composition. Here, we introduce techniques that go beyond identifying the taxa present within a sample and examine the biological function of the microbiome or the host-microbiome interaction.

Introduction

Over the past 10 years, there has been a dramatic rise in the number of microbiome studies [1]. Two key developments, which have led to the recent explosion of interest in the microbiome and studies cataloging and defining the nature of microbial communities in a variety of biological systems, are the introduction of culture-independent analysis techniques and the development of next-generation sequencing along with advances in the bioinformatics support needed to facilitate this analysis. This has allowed researchers to circumvent the need to culture bacteria for identification. This is particularly beneficial given the difficulty of culturing of both obligatory anaerobic bacteria and other bacteria with unique or as-yet-undefined growth requirements. In addition, it is beneficial to study the microbiome as a whole as many organisms are co-dependent on each other within a niche. For example, many microbes take advantage of the metabolic abilities of other microbes within a community that break down compounds that they cannot digest by themselves, or they remove metabolic by-products of synergistic bacteria, allowing

more efficient use of dietary substrates [2,3]. Studying any given microbe in isolation will undoubtedly provide a wealth of information regarding its functional capacity but will not necessarily be reflective of its role in the larger complex microbiome from which it was isolated.

The culture-independent technique that has become the "gold standard" for microbial profiling is 16S ribosomal RNA (rRNA) gene sequencing. The use of the 16S rRNA gene for identifying bacterial taxa was pioneered by Carl Woese and others in the 1980s when it was shown that phylogenetic relationships of bacteria could be determined by comparing a stable portion of the genome, with the 16S rRNA gene being one of several possible marker genes found in all bacteria and archaea [4,5]. By using universal primers to constant regions of the 16S rRNA gene, one can amplify and sequence various hypervariable regions within this gene. These hypervariable regions have a high degree of interspecies variability that can be compared with known sequences in reference databases for taxonomic identification [6]. Although the idea is simplistic and elegant, there are various decisions

that must be made during sample preparation and data analysis that can greatly influence the conclusions. For information regarding these technical decisions and how they influence the analysis as well as a detailed description of 16S rRNA sequencing and its caveats, please see [7,8].

In addition to these technical challenges (considerations), the selection of sample type will greatly affect the conclusions of your sequencing analysis. For example, studies of the gut microbiome typically rely on the sampling of stool. This is primarily due to sample quantity and the non-invasive nature of stool collection, whereas endoscopic biopsies are much more invasive and limited in quantity but may provide a more accurate picture of those microbes directly in contact with, or more likely to influence, the host. Studies have shown that there is a great deal of difference between the stool microbiome and that of the mucosa-associated microbiome obtained from biopsy samples [9,10]. Even within biopsy samples, absolute numbers and the diversity of the microbes present vary along the length of the gastrointestinal tract [9,10].

While DNA sequencing has been available since the 1970s, this traditional sequencing, such as Sanger sequencing, was very costly, had a low throughput, and took a considerable amount of time to perform. The invention of newer massively paralleled sequencing technology, commonly referred to as next-generation sequencing, has led to a remarkable increase in the number of microbiome studies. Although next-generation sequencing techniques generally offer shorter contiguous DNA sequencing reads compared with Sanger sequencing, they do allow millions of sequencing reactions to occur in parallel within a single sequencing run [11]. This has significantly driven down the cost of sequencing and will likely continue to do so as this technology advances. The impact that 16S rRNA gene sequencing paired with next-generation sequencing has had on the field of the microbiome is evident from the surge in the number of references obtained from a simple PubMed search. In 2001, there were 74 articles that cited "microbiome", whereas in 2013 there were 3,254.

Most of the studies undertaken so far have served to catalog the microbial species present in an environmental sample (for example, in the gut microbiome of healthy individuals) [12,13]. A number of studies have focused on examining microbial profiles associated with particular disease pathology. This type of cataloging study has been done extensively in subjects with inflammatory bowel disease (IBD) [14–19], in which there is strong evidence to suggest that the microbiome is a major

contributing factor to disease development [20–25]. Although all of this information has contributed significantly to our understanding of the composition of the human gut microbiome, additional work and alternative analyses are needed to understand how a given microbial community functions and interacts with host physiology, and influences health and disease.

When one looks at these cataloging studies as a whole, it is evident that there does not appear to be a "core microbiota", meaning that there is not a species or group of species that appears to be common to everyone. It is possible that different species or communities as a whole have functional redundancies. If one considers the microbiome as a complex organism, it is important to examine the functional capacity of that microbiome as it may reflect "core" functions and differentiate healthy from disease-related microbiomes. This suggests that understanding the function of the microbiome may be more important than simply cataloging the individual components (i.e. organisms). The next phase in microbiome studies will undoubtedly focus on biological function, extending beyond descriptions of the organisms that are present, to understand how a given microbiome population may function to affect the host and participate in disease processes. Here, we examine techniques that go beyond identifying the taxa present within a sample and examine the biological function of the microbiome or the host-microbiome interaction.

Metagenomics by shotgun sequencing

Metagenomics is the study of all genes contained within a community and theoretically allows assessment of the functional potential of a given microbiome, including bacterial, eukaryotic, and viral functions [26,27]. Although this is potentially a very powerful tool, it is important to keep in mind that identifying the presence of a gene with an assigned function is different from knowing whether the gene is actually expressed (transcribed and translated).

It has been suggested that there is a core functional microbiome, meaning that although there is no evidence for a set of common species shared between all individuals, there nonetheless may be common genes or pathways detected in the microbiome of all individuals [28,29]. However, certain functional pathways may be enriched because they are required for existence rather than indicating a role for these pathways in host-microbiome interactions. For example, Lozupone *et al.* [30] showed that the core functions of the gut microbiome include metabolic pathways important for survival of the organisms living in the gut environment, such as carbohydrate and amino acid metabolism.

The most common way to perform metagenomics is through the application of shotgun sequencing. In shotgun sequencing, extracted community DNA is sheared and sequenced. The resulting reads are blasted against various databases—such as the functional database: Kyoto Encyclopedia of Genes and Genomes (KEGG) or the protein database: Clusters of Orthologous Groups (COG)—in order to obtain an individual pathway or gene assignments. This provides a cataloging of the gene sequences and their corresponding COG classifications that are present in the given sample. Depending on the level of resolution (the ability to assign reads to broad metabolic functions of very specialized functions), it can be difficult to extrapolate the results into biological relevance. For example, by presenting data by COG categories, sequencing reads may fall into the COG categories, such as "J", "T", or "Q", which encode for translation, signal transduction, and secondary metabolite biosynthesis, transport, and catabolism, respectively [31]. Yet such functional categories are very broad, making it difficult to identify specific elements of a given pathway.

The main downside of metagenomic analysis is that shotgun sequencing requires extensive sequencing reads (i.e. increased depth of sequencing) to ensure sufficient coverage of the entire metagenome [32]. This is important because variable functions that could be biologically important, such as butyrate or antibiotic production, which are usually restricted to select species or strains, may be difficult to detect if sequencing depth is insufficient [30]. In addition, a large proportion of the sequences generated are unassigned because they are not annotated or are not found in current databases. This reflects the incomplete collection of curated reference genomes available for comparisons. Currently, there does not appear to be a consensus in the literature regarding the depth of sequencing required to capture all of the functions of the microbiome. It has been suggested that the amount of sequencing data required to capture all of the bacterial functions of the human gut microbiome is approximately 7 gigabytes, but estimation is difficult to perform due to its dependency on bacterial genome size, the sequencing coverage of the bacterial genome, presence of plasmids, the complexity of the community, and error rate of the sequencing technology employed [33,34].

However, metagenomic sequencing has been used to identify relevant pathways that may be worth targeting therapeutically. For example, the microbiome of subjects with IBD were shown to have decreased amino acid biosynthesis and carbohydrate metabolism compared with healthy subjects. Specifically, subjects with ileal Crohn's disease (CD) were shown to have reduced vitamin biosynthesis and increased oxidative stress compared with

ulcerative colitis (UC) or non-ileal CD [24]. These pathways could potentially be assessed in greater detail by using gene-targeting strategies or methods that rely on the direct cloning of DNA fragments extracted from uncultured microbial communities to identify and exploit novel therapeutic molecules [35,36]. The use of these more targeted approaches has already been initiated in the field of soil microbiology to attempt to identify new antibiotic-resistant genes or antimicrobials from uncultured microorganisms [36–39]. For example, in one study utilizing functional metagenomics, genes encoding antimicrobial agents were screened by using clone libraries constructed from DNA isolated from arid soil bacterial samples. After expression in *Streptomyces albus*, it was found that the recombinant bacteria showed inhibitory activity against methicillin-resistant *Staphylococcus aureus* as well as vancomycin-resistant *Enterococcus faecalis* [36].

These functional metagenomic approaches have only begun to be exploited to examine host-microbe interactions. In a two-part high-throughput screen, metagenomic libraries generated from human fecal microbiota samples were cloned into bacterial cells. Lysates from these bacterial cell suspensions were subsequently added to eukaryotic cell lines, including a human colonic cell line, to identify genes that inhibit or enhance eukaryotic cell growth [40]. Screens like this could be adopted to test a number of host outputs and help define host-microbe interactions but also serve to direct drug development.

Metagenomics inferred by imputing genetic functional capacity: PICRUSt

Owing to the inherent complexity and expense of the substantial number of sequencing reads required for functional diversity profiling using a shotgun sequencing approach, researchers have developed software that imputes function based on 16S rRNA profiles. The Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) software uses genus or species identifiers assigned from 16S rRNA gene sequencing data, to infer function based on known full-reference genomes [41]. This software takes into account several factors important for metagenomic prediction, such as the availability of pan and core genomes of microbioal reference taxa [42] and 16S rRNA copy number among bacterial taxa [8]. The software generates functional classifications based on KEGG [43] orthology and COG [44]. As such, the same caveats apply to these data as with shotgun sequencing. Outputs from KEGG and COG can often represent broad functions, making it difficult to narrow the output down to specific elements of a given pathway.

The output from the PICRUSt software shows an average of 80% correlation with shotgun sequencing [41] within

microbial communities, such as the human gut microbiome, where the number of fully sequenced genomes is greatest [45]. However, unlike shotgun sequencing, PICRUSt does not reflect strain diversity such as that which may be observed in numerous bacterial species, including pathobiont strains [46]. It also cannot yet impute viral or eukaryotic organism function. Furthermore, results from this analysis are dependent on, and thus biased by, the hypervariable region or regions (i.e. V1-V9) of 16S rRNA sequenced, as taxonomic assignment is affected by the region selected for sequencing [47]. In summary, PICRUSt serves as a "poor man's metagenomics". Although it still has room to improve, it is a promising inexpensive tool that can be applied to any study with 16S rRNA data, and continuous efforts to sequence more gut microbial genomes will improve the accuracy of the software with time [41].

Metatranscriptomics

Metatranscriptomics measures the total RNA present in a community of bacteria and, in a manner similar to metagenomics, provides a snapshot of the functional potential of a bacterial community as determined by the genes that have been activated and transcribed. The benefit of metatranscriptomics is that it has a greater likelihood of differentiating between genes that are expressed and active versus those which are not. In this way, metatranscriptomics provides not only a snapshot of the functional potential of a bacterial community but also a means of measuring the actual metabolic activity of that community. Initial research based on computational analysis of the microbial genome suggested that prokaryotic transcription was a relatively simple process. However, more recent studies demonstrate that there is a significant amount of complexity, with previously unrecognized functional RNA elements, including non-coding RNAs and riboswitches, having a more dominant role than previously recognized. Furthermore, although it was previously believed that the structure and transcription of operons were fairly static, recent studies have demonstrated the presence of alternative operon structures with increased regulatory potential [48]. Such functional complexity highlights the added benefit of using non-DNA-based methods in community analysis.

Metatranscriptomics involves extraction of the total RNA from a microbial community sample, conversion of total RNA, or enriched messenger RNA (mRNA) to complementary DNA, combined with the use of either next-generation sequencing or microarrays to determine which portions of the genome are expressed. Despite the advantages and knowledge gained from studies of the transcriptome, technical difficulties in processing this genetic material often make this approach logistically more difficult. Enrichment of mRNA is often necessary due to the large amount (>75%) of rRNA and transfer RNA present in cells [49,50] and can be difficult as the polyadenylated tail used to isolate mRNA from eukaryotic organisms does not exist in bacterial cells. Furthermore, bacterial mRNA seems to be particularly unstable, with a very short half-life, demanding the immediate and efficient use of preservatives upon sampling or immediate preparation and extraction of RNA from samples [50-52]. Resulting reads must also be mapped against known bacterial genomes in order to identify the nature and function of a given gene, a major limitation given the low number of organisms with a fully sequenced genome. Despite these challenges, several recent studies have begun to offer insight into the functional role of microbial communities in human health.

To date, metatranscript studies of the human gut microbiome have been small in scale and demonstrate a large degree of inter-individual variability. However, similar to metagenomic observations, these studies have revealed relatively high levels of transcripts corresponding to basic cellular processes. Transcripts corresponding to "RNA polymerase", "ribosome", "pyruvate metabolism", and "glycolysis" among others are detected at relatively increased abundances across samples [53,54]. Interestingly, several studies have demonstrated alterations in transcriptional responses within the microbiome of individuals exposed to different food sources [55,56]. Indeed, a large component of the metatranscriptome in stool is related to carbohydrate metabolism [53]. Most importantly, recent studies have shown that alterations in gene expression and function can occur in response to dietary or probiotic intervention, independent of alterations in specific microbial community structure composition, further highlighting the importance of such functional studies in understanding the true impact of environmental factors on the host microbiome [57].

Metabolomics

Metabolomics is the study of complex biological samples, which aims to quantify and identify small molecules that are the by-products of metabolism. The term "metabolic profile" is often used to describe the collection of metabolites found within a biological sample, much like the term microbiota is used to encompass all of the bacteria present within a given environmental sample. Both genetics and environment greatly alter metabolism, but metabolomics allows for monitoring of the outcome. Like the host, the microbiome produces metabolites that give us an idea of its function and how it may interact with the host through host-microbe co-metabolism.

Metabolic profiles have already proven to be useful in distinguishing between healthy and diseased individuals, thus providing biomarkers of disease. One example would be the use of metabolomics in IBD. Several studies have identified not only metabolites that distinguish healthy individuals from those with IBD [58-61] but also metabolites that distinguish between UC and CD [59,61]. The benefits of developing such biomarkers based on blood, urine, or stool are that these biomarkers provide a non-invasive tool for disease identification/phenotyping, mechanistic insight into disease processes, insight into drug metabolism, including drug toxicity, and possible insight into monitoring therapeutic interventions and may predict an individual's response to therapy.

The "targeted" approach to metabolomics can measure a specific metabolite or class of metabolites (e.g. bile acids). An "untargeted" metabolomic approach provides a broad cataloging of many different classes of metabolites involving a broad range of metabolic pathways. This can be achieved by employing one or more of several analytical techniques available; however, it is important to note that no single analytical technique can be used to measure all types of metabolites. The most common techniques for metabolic profiling are mass spectrometry (MS) and nuclear magnetic resonance (NMR). Both of these techniques can be used to identify metabolic profiles in a variety of biological samples, including urine, serum, fecal extracts, and biopsy samples (e.g. biopsied colon tissue). MS is generally paired with an upstream separation technique, such as gas chromatography (GC-MS) or liquid chromatography (LC-MS), and works by distinguishing molecules based on their mass-to-charge ratios and retention times. NMR, on the other hand, measures the absorption of electromagnetic radiation by different nuclei within a compound when it is placed under a magnetic field. Unlike MS, NMR is a non-destructive detection method that allows the sample to be recovered for further analysis. Additional benefits are that NMR often requires only small sample amounts, on the order of milligram or sub-milligram levels, and requires little to no sample preparation. For a more comprehensive review of metabolomics analytical techniques, please see Lindon *et al.* [62]. The data analysis also presents a major challenge when dealing with these analytical methods. Complex datasets are generated from these technologies and their interpretation requires the use of multivariate statistical analysis.

Studies of germ-free versus conventional animals (those which harbor a traditional microbiome) have shown that the microbiome is responsible for producing a number of metabolites capable of interacting with the host [63,64]. Importantly, the microbiome is responsible for producing short-chain fatty acids which are utilized by the host (e.g. serving as an important energy source for colonic epithelial cells) [65,66].

A series of elegant studies by the Hazen lab [67–69] has shown us how the use of metabolomics has identified a role of the microbiome linked with diet in cardiovascular disease. The initial study by Wang *et al.* [67] set out to identify small molecules associated with individuals with cardiovascular disease. This untargeted blood screen found that the dietary metabolite phosphatidylcholine was converted by the gut microbiome to trimethylamine N-oxide (TMAO), which by a yet unknown mechanism contributes to atherosclerosis. Follow-up studies by this group utilizing both well-delineated human and animal studies show that the use of antibiotics suppressed the increase in plasma levels of TMAO when subjects were given a dietary phosphatidylcholine challenge [68] or L-carnitine challenge [69], providing further evidence that the production of TMAO is catalyzed by intestinal microbial metabolism. This is an excellent example of how the use of metabolomics can provide mechanistic insight into the role of the microbiome in disease pathogenesis and further generate new areas of study.

Future of microbiome studies

Although all of the -omics techniques discussed here are culture-independent, it continues to be important to study microbes in isolation to fully grasp their functional potential. Laboratory techniques have continued to progress, allowing more microbes than ever to be cultured [70]. By isolating these microbes, we can not only interrogate their function but genetically manipulate them by using insertion sequencing [71] or other techniques to determine which genes are vital to various processes (i.e. colonization of the gut, growth with certain dietary components, competing with various pathogens).

Employing -omics-based approaches will provide an understanding of the role of the microbiome in host health and disease. However, to determine whether the microbiome is causal in disease pathogenesis, integration of these techniques with *in vitro* and *in vivo* studies will be required. Perhaps some of the most useful *in vitro* techniques include bioreactor systems that mimic the conditions found in the gastrointestinal tract. This allows culturing of bacteria within a complex microbial community and a complex environment that more closely resembles the gut. For a review of a variety of *in vitro* techniques available, refer to reference [72].

Undoubtedly, the use of germ-free animals not only has been useful for determining the function of the microbiome in immune development and host physiology [73] but has also proved to be a powerful tool to show causality of the microbiome in disease. By colonizing germ-free animals with defined or whole microbial communities, researchers are able to limit confounding variables, such as diet and environment, to isolate the contribution of the microbiome to pathologies such as obesity, diabetes, and cardiovascular disease to name a few [74,75]. More recent gnotobiotic studies (germ-free or formerly germ-free prior to introduction of microbes) have utilized mice; however, there has also been an increase in the use of germ-free zebrafish because of the ease of working with this model and its relatively inexpensive cost and an increase in the use of germ-free pigs, as the immune system of this animal more closely resembles that of humans. Each model has its own benefits and disadvantages. For a review of gnotobiotic animal models in microbiome research, please see references [72,76].

Conclusions

The development and application of new technologies involved in next-generation sequencing have led to an explosion of investigative work focused on the characterization of the microbiome in health and disease. This has certainly supported the pre-existing notion that the gut microbiome is important in several gastrointestinal diseases and has suggested a role for microbes in this niche in promoting diseases affecting other organs, such as cardiovascular disease, diabetes, and cancer [77,78]. Although we are still learning how to more accurately identify which bacteria are present and have begun to expand this inventory to include viruses and fungi, we must also go beyond cataloging and examine the functional role that the microbiome plays in health and disease. This will require the use of well-designed experiments paired with appropriate techniques, such as metagenomics, metatranscriptomics, and metabolomics, evaluated by more sophisticated bioinformatic tools to assess the biological importance and mechanisms of the microbiome and host-microbiome interactions.

Abbreviations

CD, Crohn's disease; COG, Clusters of Orthologous Groups; IBD, inflammatory bowel disease; KEGG, Kyoto Encyclopedia of Genes and Genomes; mRNA, messenger RNA; MS, mass spectrometry; NMR, nuclear magnetic resonance; PICRUSt, Phylogenetic Investigation of Communities by Reconstruction of Unobserved States; rRNA, ribosomal RNA; TMAO, trimethylamine N-oxide; UC, ulcerative colitis.

Disclosures

The authors declare that they have no disclosures.

Acknowledgments

References

1. Jones S: **Trends in microbiome research.** *Nat Biotechnol* 2013, **31**:277.

2. Rey FE, Gonzalez MD, Cheng J, Wu M, Ahern PP, Gordon JI: **Metabolic niche of a prominent sulfate-reducing human gut bacterium.** *Proc Natl Acad Sci U S A* 2013, **110**:13582-7.

3. Stams Alfons J M, Plugge CM: **Electron transfer in syntrophic communities of anaerobic bacteria and archaea.** *Nat Rev Microbiol* 2009, **7**:568-77.

4. Woese CR: **Bacterial evolution.** *Microbiol Rev* 1987, **51**:221-71.

5. Woese CR, Fox GE: **Phylogenetic structure of the prokaryotic domain: the primary kingdoms.** *Proc Natl Acad Sci U S A* 1977, **74**:5088-90.

6. Weisburg WG, Barns SM, Pelletier DA, Lane DJ: **16S ribosomal DNA amplification for phylogenetic study.** *J Bacteriol* 1991, **173**:697-703.

7. Tyler AD, Smith MI, Silverberg MS: **Analyzing the Human Microbiome: A 'How To' guide for Physicians.** *Am J Gastroenterol* 2014.

8. Větrovský T, Baldrian P: **The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses.** *PLoS ONE* 2013, **8**:e57923.

9. Peterson DA, Frank DN, Pace NR, Gordon JI: **Metagenomic approaches for defining the pathogenesis of inflammatory bowel diseases.** *Cell Host Microbe* 2008, **3**:417-27.

    F1000Prime RECOMMENDED

10. Stearns JC, Lynch Michael D J, Senadheera DB, Tenenbaum HC, Goldberg MB, Cvitkovitch DG, Croitoru K, Moreno-Hagelsieb G, Neufeld JD: **Bacterial biogeography of the human digestive tract.** *Sci Rep* 2011, **1**:170.

11. Mardis ER: **Next-generation sequencing platforms.** *Annu Rev Anal Chem (Palo Alto Calif)* 2013, **6**:287-303.

12. Nelson KE, Weinstock GM, Highlander SK, Worley KC, Creasy HH, Wortman JR, Rusch DB, Mitreva M, Sodergren E, Chinwalla AT, Feldgarden M, Gevers D, Haas BJ, Madupu R, Ward DV, Birren BW, Gibbs RA, Methe B, Petrosino JF, Strausberg RL, Sutton GG, White OR, Wilson RK, Durkin S, Giglio MG, Gujja S, Howarth C, Kodira CD, Kyrpides N, Mehta T, *et al.*: **A catalog of reference genomes from the human microbiome.** *Science* 2010, **328**:994-9.

    F1000Prime RECOMMENDED

13. Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, Baker CC, Di Francesco V, Howcroft TK, Karp RW, Lunsford RD, Wellington CR, Belachew T, Wright M, Giblin C, David H, Mills M, Salomon R, Mullins C, Akolkar B, Begg L, Davis C, Grandison L, Humble M, Khalsa J, Little AR, *et al.*: **The NIH Human Microbiome Project.** *Genome Res* 2009, **19**:2317-23.

    F1000Prime RECOMMENDED

14. Manichanh C, Borruel N, Casellas F, Guarner F: **The gut microbiota in IBD.** *Nat Rev Gastroenterol Hepatol* 2012, **9**:599-608.

15. Bibiloni R, Mangold M, Madsen KL, Fedorak RN, Tannock GW: **The bacteriology of biopsies differs between newly diagnosed, untreated, Crohn's disease and ulcerative colitis patients.** *J Med Microbiol* 2006, **55**:1141-9.

16. Willing B, Halfvarson J, Dicksved J, Rosenquist M, Järnerot G, Engstrand L, Tysk C, Jansson JK: **Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease.** *Inflamm Bowel Dis* 2009, **15**:653-60.

17. Willing BP, Dicksved J, Halfvarson J, Andersson AF, Lucio M, Zheng Z, Järnerot G, Tysk C, Jansson JK, Engstrand L: **A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes.** *Gastroenterology* 2010, **139**:1844-1854.e1.

18. Lepage P, Häsler R, Spehlmann ME, Rehman A, Zvirbliene A, Begun A, Ott S, Kupcinskas L, Doré J, Raedler A, Schreiber S: **Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis.** *Gastroenterology* 2011, **141**:227-36.

19. Wu GD, Bushmanc FD, Lewis JD: **Diet, the human gut microbiota, and IBD.** *Anaerobe* 2013, **24**:117-20.

20. Steinhart AH, Feagan BG, Wong CJ, Vandervoort M, Mikolainis S, Croitoru K, Seidman E, Leddin DJ, Bitton A, Drouin E, Cohen A, Greenberg GR: **Combined budesonide and antibiotic therapy for active Crohn's disease: a randomized controlled trial.** *Gastroenterology* 2002, **123**:33-40.

21. D'Haens GR, Geboes K, Peeters M, Baert F, Penninckx F, Rutgeerts P: **Early lesions of recurrent Crohn's disease caused by infusion of intestinal contents in excluded ileum.** *Gastroenterology* 1998, **114**:262-7.

22. Petnicki-Ocwieja T, Hrncir T, Liu Y, Biswas A, Hudcovic T, Tlaskalova-Hogenova H, Kobayashi KS: **Nod2 is required for the regulation of commensal microbiota in the intestine.** *Proc Natl Acad Sci U S A* 2009, **106**:15813-8.

    F1000Prime RECOMMENDED

23. Balfour Sartor R: **Enteric Microflora in IBD: Pathogens or Commensals?** *Inflamm Bowel Dis* 1997, **3**:230-5.

24. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C: **Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment.** *Genome Biol* 2012, **13**:R79.

    F1000Prime RECOMMENDED

25. Papa E, Docktor M, Smillie C, Weber S, Preheim SP, Gevers D, Giannoukos G, Ciulla D, Tabbaa D, Ingram J, Schauer DB, Ward DV, Korzenik JR, Xavier RJ, Bousvaros A, Alm EJ: **Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease.** *PLoS ONE* 2012, **7**:e39242.

26. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF: **Community structure and metabolism through reconstruction of microbial genomes from the environment.** *Nature* 2004, **428**:37-43.

    F1000Prime RECOMMENDED

27. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers Y, Smith HO: **Environmental genome shotgun sequencing of the Sargasso Sea.** *Science* 2004, **304**:66-74.

    F1000Prime RECOMMENDED

28. Turnbaugh PJ, Gordon JI: **The core gut microbiome, energy balance and obesity.** *J Physiol (Lond.)* 2009, **587**:4153-8.

29. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI: **A core gut microbiome in obese and lean twins.** *Nature* 2009, **457**:480-4.

    F1000Prime RECOMMENDED

30. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R: **Diversity, stability and resilience of the human gut microbiota.** *Nature* 2012, **489**:220-30.

    F1000Prime RECOMMENDED

31. Tatusov RL, Galperin MY, Natale DA, Koonin EV: **The COG database: a tool for genome-scale analysis of protein functions and evolution.** *Nucleic Acids Res* 2000, **28**:33-6.

32. Knight R, Jansson J, Field D, Fierer N, Desai N, Fuhrman JA, Hugenholtz P, van der Lelie, Daniel, Meyer F, Stevens R, Bailey MJ, Gordon JI, Kowalchuk GA, Gilbert JA: **Unlocking the potential of metagenomics through replicated experimental design.** *Nat Biotechnol* 2012, **30**:513-20.

33. Ni J, Yan Q, Yu Y: **How much metagenomic sequencing is enough to achieve a given goal?** *Sci Rep* 2013, **3**:1968.

34. Kunin V, Copeland A, Lapidus A, Mavromatis K, Hugenholtz P: **A bioinformatician's guide to metagenomics.** *Microbiol Mol Biol Rev* 2008, **72**:557-78, Table of Contents.

35. Culligan EP, Sleator RD, Marchesi JR, Hill C: **Metagenomics and novel gene discovery: Promise and potential for novel therapeutics.** *Virulence* 2014, **5**:399-412.

36. Feng Z, Kallifidas D, Brady SF: **Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites.** *Proc Natl Acad Sci USA* 2011, **108**:12629-34.

37. Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J, Goodman RM: **Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms.** *Appl Environ Microbiol* 2000, **66**:2541-7.

38. Riesenfeld CS, Goodman RM, Handelsman J: **Uncultured soil bacteria are a reservoir of new antibiotic resistance genes.** *Environ Microbiol* 2004, **6**:981-9.

    F1000Prime RECOMMENDED

39. Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer Morten O A, Dantas G: **The shared antibiotic resistome of soil bacteria and human pathogens.** *Science* 2012, **337**:1107-11.

    F1000Prime RECOMMENDED

40. Gloux K, Leclerc M, Iliozer H, L'Haridon R, Manichanh C, Corthier G, Nalin R, Blottière HM, Doré J: **Development of high-throughput phenotyping of metagenomic clones from the human gut microbiome for modulation of eukaryotic cell growth.** *Appl Environ Microbiol* 2007, **73**:3734-7.

    F1000Prime RECOMMENDED

41. Langille Morgan G I, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber, Rebecca L, Knight R, Beiko RG, Huttenhower C: **Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences.** *Nat Biotechnol* 2013, **31**:814-21.

    F1000Prime RECOMMENDED

42. Collins RE, Higgs PG: **Testing the infinitely many genes model for the evolution of the bacterial core genome and pangenome.** *Mol Biol Evol* 2012, **29**:3413-25.

43. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M: **KEGG for integration and interpretation of large-scale molecular data sets.** *Nucleic Acids Res* 2012, **40**:D109-14.

44. Tatusov RL, Koonin EV, Lipman DJ: **A genomic perspective on protein families.** *Science* 1997, **278**:631-7.

45. Markowitz VM, Chen IA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC: **IMG: the Integrated Microbial Genomes database and comparative analysis system.** *Nucleic Acids Res* 2012, **40**:D115-22.

46. Ferrario C, Ricci G, Milani C, Lugli GA, Ventura M, Eraclio G, Borgo F, Fortina MG: **Lactococcus garvieae: where is it from? A first approach to explore the evolutionary history of this emerging pathogen.** *PLoS ONE* 2013, **8**:e84796.
F1000Prime RECOMMENDED

47. **Evaluation of 16S rDNA-based community profiling for human microbiome research.** *PLoS ONE* 2012, **7**:e39315.

48. Sorek R, Cossart P: **Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity.** *Nat Rev Genet* 2010, **11**:9-16.
F1000Prime RECOMMENDED

49. Passalacqua KD, Varadarajan A, Ondov BD, Okou DT, Zwick ME, Bergman NH: **Structure and complexity of a bacterial transcriptome.** *J Bacteriol* 2009, **191**:3203-11.
F1000Prime RECOMMENDED

50. Condon C: **Maturation and degradation of RNA in bacteria.** *Curr Opin Microbiol* 2007, **10**:271-8.

51. Deutscher MP: **Degradation of stable RNA in bacteria.** *J Biol Chem* 2003, **278**:45041-4.

52. van Vliet Arnoud H M: **Next generation sequencing of microbial transcriptomes: challenges and opportunities.** *FEMS Microbiol Lett* 2010, **302**:1-7.

53. Booijink Carien C G M, Boekhorst J, Zoetendal EG, Smidt H, Kleerebezem M, de Vos Willem M: **Metatranscriptome analysis of the human fecal microbiota reveals subject-specific expression profiles, with genes encoding proteins involved in carbohydrate metabolism being dominantly expressed.** *Appl Environ Microbiol* 2010, **76**:5533-40.
F1000Prime RECOMMENDED

54. Turnbaugh PJ, Quince C, Faith JJ, McHardy AC, Yatsunenko T, Niazi F, Affourtit J, Egholm M, Henrissat B, Knight R, Gordon JI: **Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins.** *Proc Natl Acad Sci U S A* 2010, **107**:7503-8.
F1000Prime RECOMMENDED

55. González R, Klaassens ES, Malinen E, de Vos Willem M, Vaughan EE: **Differential transcriptional response of Bifidobacterium longum to human milk, formula milk, and galactooligosaccharide.** *Appl Environ Microbiol* 2008, **74**:4686-94.
F1000Prime RECOMMENDED

56. O'Connell Motherway M, Zomer A, Leahy SC, Reunanen J, Bottacini F, Claesson MJ, O'Brien F, Flynn K, Casey PG, Munoz Jose Antonio Moreno, Kearney B, Houston AM, O'Mahony C, Des Higgins G, Shanahan F, Palva A, de Vos Willem M, Fitzgerald GF, Ventura M, O'Toole PW, van Sinderen D: **Functional genome analysis of Bifidobacterium breve UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor.** *Proc Natl Acad Sci USA* 2011, **108**:11217-22.
F1000Prime RECOMMENDED

57. McNulty NP, Yatsunenko T, Hsiao A, Faith JJ, Muegge BD, Goodman AL, Henrissat B, Oozeer R, Cools-Portier S, Gobert G, Chervaux C, Knights D, Lozupone CA, Knight R, Duncan AE, Bain JR, Muehlbauer MJ, Newgard CB, Heath AC, Gordon JI: **The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins.** *Sci Transl Med* 2011, **3**:106ra106.
F1000Prime RECOMMENDED

58. Lin H, Helsby NA, Rowan DD, Ferguson LR: **Using metabolomic analysis to understand inflammatory bowel diseases.** *Inflamm Bowel Dis* 2011, **17**:1021-9.

59. Schicho R, Shaykhutdinov R, Ngo J, Nazyrova A, Schneider C, Panaccione R, Kaplan GG, Vogel HJ, Storr M: **Quantitative metabolomic profiling of serum, plasma, and urine by (1)H NMR spectroscopy discriminates between patients with inflammatory bowel disease and healthy individuals.** *J Proteome Res* 2012, **11**:3344-57.
F1000Prime RECOMMENDED

60. Lu K, Knutson CG, Wishnok JS, Fox JG, Tannenbaum SR: **Serum metabolomics in a Helicobacter hepaticus mouse model of inflammatory bowel disease reveal important changes in the microbiome, serum peptides, and intermediary metabolism.** *J Proteome Res* 2012, **11**:4916-26.

61. Hisamatsu T, Okamoto S, Hashimoto M, Muramatsu T, Andou A, Uo M, Kitazume MT, Matsuoka K, Yajima T, Inoue N, Kanai T, Ogata H, Iwao Y, Yamakado M, Sakai R, Ono N, Ando T, Suzuki M, Hibi T: **Novel, objective, multivariate biomarkers composed of plasma amino acid profiles for the diagnosis and assessment of inflammatory bowel disease.** *PLoS ONE* 2012, **7**:e31131.
F1000Prime RECOMMENDED

62. Lindon JC, Holmes E, Nicholson JK: **Metabonomics: systems biology in pharmaceutical research and development.** *Curr Opin Mol Ther* 2004, **6**:265-72.

63. Holmes E, Li JV, Athanasiou T, Ashrafian H, Nicholson JK: **Understanding the role of gut microbiome-host metabolic signal disruption in health and disease.** *Trends Microbiol* 2011, **19**:349-59.

64. Martin FJ, Dumas M, Wang Y, Legido-Quigley C, Yap, Ivan K S, Tang H, Zirah S, Murphy GM, Cloarec O, Lindon JC, Sprenger N, Fay LB, Kochhar S, van Bladeren P, Holmes E, Nicholson JK: **A top-down systems biology view of microbiome-mammalian metabolic interactions in a mouse model.** *Mol Syst Biol* 2007, **3**:112.
F1000Prime RECOMMENDED

65. Roediger WE: **Role of anaerobic bacteria in the metabolic welfare of the colonic mucosa in man.** *Gut* 1980, **21**:793-8.

66. Donohoe DR, Garge N, Zhang X, Sun W, O'Connell TM, Bunger MK, Bultman SJ: **The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon.** *Cell Metab* 2011, **13**:517-26.

67. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung Y, Wu Y, Schauer P, Smith JD, Allayee H, Tang, W H Wilson, DiDonato JA, Lusis AJ, Hazen SL: **Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease.** *Nature* 2011, **472**:57-63.
F1000Prime RECOMMENDED

68. Tang WHW, Wang Z, Levison BS, Koeth RA, Britt EB, Fu X, Wu Y, Hazen SL: **Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk.** *N Engl J Med* 2013, **368**:1575-84.
F1000Prime RECOMMENDED

69. Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, Britt EB, Fu X, Wu Y, Li L, Smith JD, DiDonato JA, Chen J, Li H, Wu GD, Lewis JD, Warrier M, Brown JM, Krauss RM, Tang W H Wilson,

Bushman FD, Lusis AJ, Hazen SL: **Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis.** *Nat Med* 2013, **19**:576-85.

F1000Prime RECOMMENDED

70. Goodman AL, Kallstrom G, Faith JJ, Reyes A, Moore A, Dantas G, Gordon JI: **Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice.** *Proc Natl Acad Sci U S A* 2011, **108**:6252-7.

71. Goodman AL, Wu M, Gordon JI: **Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries.** *Nat Protoc* 2011, **6**:1969-80.

F1000Prime RECOMMENDED

72. Fritz JV, Desai MS, Shah P, Schneider JG, Wilmes P: **From meta-omics to causality: experimental models for human microbiome research.** *Microbiome* 2013, **1**:14.

73. Smith K, McCoy KD, Macpherson AJ: **Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota.** *Semin Immunol* 2007, **19**:59-69.

74. Devaraj S, Hemarajata P, Versalovic J: **The human gut microbiome and body metabolism: implications for obesity and diabetes.** *Clin Chem* 2013, **59**:617-28.

75. Brown JM, Hazen SL: **Metaorganismal nutrient metabolism as a basis of cardiovascular disease.** *Curr Opin Lipidol* 2014, **25**:48-53.

76. Kostic AD, Howitt MR, Garrett WS: **Exploring host-microbiota interactions in animal models and humans.** *Genes Dev* 2013, **27**:701-18.

77. Sekirov I, Russell SL, Antunes, L Caetano M, Finlay BB: **Gut microbiota in health and disease.** *Physiol Rev* 2010, **90**:859-904.

78. Neu J, Lorca G, Kingma Sandra D K, Triplett EW: **The intestinal microbiome: relationship to type 1 diabetes.** *Endocrinol Metab Clin North Am* 2010, **39**:563-71.