Microbiome Primer

Elaine Vo, PhD, Zain Kassam, MD MPH FRCPC, Mark Smith, PhD

Purpose: This document aims to provide a brief review of the salient literature to inform study design for researchers interested in the microbiome in general and fecal transplantation in particular.

Motive: Human microbiome research has become a topic of great interest among clinical and basic science researchers because of the strong associations between the composition of this microbial ecosystem and human health. The opportunity to understand these associations and potentially improve health outcomes through manipulations such as fecal microbiota transplantation is a particularly intriguing prospect. Catalyzing this research is a core component of OpenBiome's mission.

Introduction to the Human Microbiome

In adults, bacterial cells are known to outnumber human cells by a factor of 10, and the cumulative number of bacterial genes exceed those in the human genome by a factor of 100.[2] Both gut microbial density and diversity increase from the stomach (10 microbial cells per gram content) to the colon (10^{12} cells per gram) as well as from the epithelial surfaces to the lumen.[4] "Normal" gut microbiota are comprised predominantly of anaerobic bacteria, with a total of 500-1000 species belonging to only a few bacterial phyla. Firmicutes and Bacteroidetes are the most abundant phyla in the gut whereas Proteobacteria, Actinobacteria, Fusobacteria, Verrumicrobia, and Cyanobacteria are observed at low abundance. Colonization of the gut begins at birth, and the process of microbial community assembly is shaped by many factors including diet,[5] environmental conditions,[6] antibiotic use,[7] and host genetics.[8]

16S rRNA as a biomarker

The 16S rRNA gene is favored as a molecular marker for interrogating bacterial communities due to its nearly universal distribution among bacteria coupled with a divergence that provides useful phylogenetic information about its bacterial host. The 16S rRNA gene is highly conserved, which affords effective target sites for binding of universal primers in polymerase chain reactions (PCRs) to generate multiple copies for sequencing. Yet, the gene also contains nine hypervariable regions across its full length, which supports species-specific differentiation. The development of low-cost, high-throughput sequencing technologies has enabled researchers to rapidly generate and analyze hundreds of thousands of 16S rRNA sequence reads from a single sample, providing a window into the relative abundances of bacteria present in a given sample. This analysis can provide a valuable **secondary endpoint** for clinical trials and also enable the discovery of hidden subpopulations enrolled in studies that may

Rev. June 2016
www.openbiome.org

200 Inner Belt Road
Somerville, MA 02143

Phone/Fax: 617-575-2201
info@openbiome.org

respond differently to fecal microbiota transplantation or other microbiome-based therapies

Sampling and Sequencing Logistics

For clinicians interested in examining the microbiome in clinical study, the following best practices should be considered during study design:

1. Research on human subjects must have an approved Institutional Review Board (IRB) application filed with every study site.
2. Samples can be stored in a variety of fixative agents to preserve community composition. OpenBiome can provide collection kits for patients with RNAlater which allows samples to be stored at room temperature for up to one week after collection, facilitating at home collection as needed. Alternatively, samples can be stored in glycerol or ethanol solutions if frozen immediately. A single fixative should be used for an entire study.
3. Samples should be aliquoted and frozen at -80°C as soon as possible after collection (up to one week if stored in RNAlater). Freeze-thaw cycles should be eliminated. Once a sample is frozen, it should remain frozen until sequencing.
4. There are significant batch effects when sequencing is performed. As a result, it is strongly preferable to *collect all samples for an entire study and sequence them once* at the conclusion of the study in a single batch. This avoids the confounding of batch-effects in downstream analysis.
5. Stool contains an enormous amount of bacterial DNA. A single gram suspended in 5 ml of buffer is more than sufficient for most applications.
6. Although there are a few standard analyses that most groups perform with 16S rRNA data (such as comparison of diversity and similarity metrics) we strongly encourage you to work closely with a bioinformatician to think carefully and creatively about other analyses that you can use to extract clinically informative insights from your data.

For researchers conducting the sequencing, we recommend the following methodological controls:

1. Perform all laboratory work in an irradiated, laminar flow hood and wear personal protective equipment to prevent sample contamination.
2. Incorporate both chemical and mechanical lysis into DNA extraction protocol to ensure maximal sample lysis.
3. Perform triplicate PCR reactions per sample to minimize amplification bias, and include a no-template negative control with each PCR plate to detect contamination. Rerun PCRs until all negative controls are clear.
4. Use error-correcting barcodes attached to the 5' end of the primers to label sequences during PCR and enable subsequent tracing of all sequences back to sample identities. Error-correction based on a coding algorithm (e.g., Golay or

Hamming) permits assignment of a sequence to sample origin even if a single substitution error were to occur in the barcode during sequencing.

5. Sequence all samples together in the same run to prevent the introduction of systematic bias between sequencing runs or platforms.
6. For many sequencing platforms, the tail of sequence reads often exhibit poorer quality than the beginning of the read. Design primers to produce amplicons of such length that after paired-end sequencing, the tails of both reads overlap, thereby permitting the overlapped region to be used in quality checking (i.e., the overlapped region should be identical).

For bioinformaticians involved in sequence analysis, we recommend the following basic workflow as a starting point for molecular analysis:
1. Exclude reads that do not pass quality filter thresholds specific to the sequencing platform.
2. Trim adapter sequences.
3. Error-correct barcodes and discard sequences with irrecoverable or unrecognizable barcodes.
4. Exclude reads that do not contain primer sequences.
5. Trim barcode and primer sequences.
6. Trim low quality bases at the tail (3') end of the sequence reads.
7. Exclude reads that are shorter than a minimum length.
8. Exclude sequences that fail to overlap or contain mismatches in the overlapped region during merging of paired-end reads.
9. Detect and exclude reference-based and de novo chimeras.
10. Exclude sequences that are only observed once in the entire dataset.
11. Exclude operational taxonomic units (OTUs) that are observed only once in the entire dataset.
12. Depending on the research aims, exclude OTUs that match to chloroplast or mitochondria.
13. OTU picking and taxonomic assignment can be performed as a closed process using a reference database or as an open process incorporating both reference-based and de novo OTU classification.
14. Rarefaction analyses inform whether the achieved sequencing depth was sufficient to recover the majority of the microbial diversity in the original samples.
15. Data visualization can be done using ordination or hierarchical clustering.
16. Statistical testing identifies significant differences between putative clusters.
17. Where microbial alpha diversity indices are of interest (e.g., Shannon's diversity index), generalized linear or additive mixed models evaluate the relative significance of various fixed and random effects in explaining variation in the outcomes data.

18. Where microbial beta diversity indices are of interest (e.g., distance matrices), Partial Mantel and Multiple Matrix Regression with Randomization[46] tests determine the relative significance of various fixed and random effects in explaining variation in the outcomes data.

19. Machine learning algorithms such as Random Forests can be trained to identify classes (e.g., healthy vs diseased) based on distinct features (e.g., relative abundance of Lactobaccili). Trained models can subsequently be applied to new data in a discriminant analysis to suggest sample classification (e.g., healthy vs diseased).

Diseases associated with the microbiome

As human microbiome research has flourished under methodological advances,[14] arguably the most exciting findings have centered on the extensive impacts that microbiota can have on host health and disease. Microbiota mediate a number of human physiological processes including metabolism,[15] gut development,[16] adiposity,[17] behavior,[18] and immune function.[19] As a result, gut microbiota are implicated in the pathogenesis of many diseases including metabolic syndrome,[20] type 2 diabetes,[21] food allergies,[22] inflammatory bowel disease,[23] cancer,[24] and several more spanning both local and systemic levels.

The use of germ-free mice in microbial colonization experiments has been especially powerful in determining the transmissibility of putative, microbial-mediated phenotypes, thereby providing insights into causal relationships. From these studies, microbiota are known to decrease bone-mass density,[25] enhance postnatal gut angiogenesis,[26] and support development of robust immune responses to pathogen infection.[27] Microbial production of short-chain fatty acids from dietary carbohydrates modulates energy availability,[28] water absorption,[29] gut motility and wound healing,[30] as well as fat storage.[31,32] Furthermore, obesity is transmissible through transplant of gut microbiota from obese mice to germ-free mice, and leanness is inducible through co-housing and sharing of microbiota between lean and obese mice on the same diet.[33] Thus, microbiota are increasingly recognized as pervasive symbionts with significant capacity to impact host growth, development, and condition.

Dysbiosis of the gut microbiota refers to imbalances in gut microbiology, including deviations from the general species richness, relative abundance, and functional capacity observed in "healthy" individuals. Patients diagnosed with recurrent *Clostridum difficile* infections (CDI),[34] inflammatory bowel disease,[35] irritable bowel syndrome,[36] celiac Disease,[37] alcoholic liver disease,[38] and a growing number of other disorders being studied exhibit dysbiotic gut microbiota. These patterns have prompted investigations into modulation of the microbiota as a therapeutic avenue, and efforts to do so have encompassed dietary changes; supplementation with prebiotics, probiotics, and synbiotics; as well as fecal microbiota transplantation.

References

1. Ley, R. E. *et al.* Evolution of Mammals and Their Gut Microbes. *Science* **320,** 1647–1651 (2008).
2. Gill, S. R. *et al.* Metagenomic Analysis of the Human Distal Gut Microbiome. *Science* **312,** 1355–1359 (2006).
3. Savage, D. C. Microbial Ecology of the Gastrointestinal Tract. *Annual Review of Microbiology* **31,** 107–133 (1977).
4. Sekirov, I., Russell, S. L., Antunes, L. C. M. & Finlay, B. B. Gut Microbiota in Health and Disease. *Physiological Reviews* **90,** 859–904 (2010).
5. Turnbaugh, P. J. *et al.* The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. *Sci Transl Med* **1,** 6ra14–6ra14 (2009).
6. Spor, A., Koren, O. & Ley, R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Micro* **9,** 279–290 (2011).
7. Dethlefsen, L., Huse, S., Sogin, M. L. & Relman, D. A. The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. *PLoS Biol* **6,** e280 (2008).
8. Benson, A. K. *et al.* Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *PNAS* **107,** 18933–18938 (2010).
9. Fox, G. E. *et al.* The phylogeny of prokaryotes. *Science* **209,** 457–463 (1980).
10. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *PNAS* **74,** 5463–5467 (1977).
11. Prober, J. M. *et al.* A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238,** 336–341 (1987).
12. Schuster, S. C. Next-generation sequencing transforms today's biology. *Nat. Methods* **5,** 16–18 (2008).
13. Lamendella, R., VerBerkmoes, N. & Jansson, J. K. 'Omics' of the mammalian gut – new insights into function. *Current Opinion in Biotechnology* **23,** 491–500 (2012).
14. Weinstock, G. M. Genomic approaches to studying the human microbiota. *Nature* **489,** 250–256 (2012).
15. Tremaroli, V. & Bäckhed, F. Functional interactions between the gut microbiota and host metabolism. *Nature* **489,** 242–249 (2012).
16. Sommer, F. & Bäckhed, F. The gut microbiota — masters of host development and physiology. *Nat Rev Micro* **11,** 227–238 (2013).
17. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–484 (2009).
18. Heijtz, R. D. *et al.* Normal gut microbiota modulates brain development and behavior. *PNAS* **108,** 3047–3052 (2011).
19. Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions Between the Microbiota and the Immune System. *Science* **336,** 1268–1273 (2012).
20. Tilg, H. Obesity, metabolic syndrome, and microbiota: multiple interactions. *J. Clin. Gastroenterol.* **44 Suppl 1,** S16–18 (2010).
21. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490,** 55–60 (2012).
22. Noval Rivas, M. *et al.* A microbiota signature associated with experimental food allergy promotes allergic sensitization and anaphylaxis. *Journal of Allergy and Clinical Immunology* **131,** 201–212 (2013).
23. Manichanh, C., Borruel, N., Casellas, F. & Guarner, F. The gut microbiota in IBD. *Nat Rev Gastroenterol Hepatol* **9,** 599–608 (2012).
24. Kostic, A. D. *et al.* Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res.* **22,** 292–298 (2012).
25. Sjögren, K. *et al.* The gut microbiota regulates bone mass in mice. *Journal of Bone and Mineral Research* **27,** 1357–1367 (2012).

26. Stappenbeck, T. S., Hooper, L. V. & Gordon, J. I. Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *PNAS* **99,** 15451–15455 (2002).
27. Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9,** 313–323 (2009).
28. Bergman, E. N. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiological Reviews* **70,** 567–590 (1990).
29. Hume, I. D. in *Gastrointestinal Microbiology* (eds. Mackie, R. I. & White, B. A.) 84–115 (Springer US, 1997). at <http://link.springer.com/chapter/10.1007/978-1-4615-4111-0_4>
30. Scheppach, W. Effects of short chain fatty acids on gut morphology and function. *Gut* **35,** S35–S38 (1994).
31. Wostmann, B. S., Larkin, C., Moriarty, A. & Bruckner-Kardoss, E. Dietary intake, energy metabolism, and excretory losses of adult male germfree Wistar rats. *Lab Anim Sci* **33,** 46–50 (1983).
32. Bäckhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *PNAS* **101,** 15718–15723 (2004).
33. Ridaura, V. K. *et al.* Gut Microbiota from Twins Discordant for Obesity Modulate Metabolism in Mice. *Science* **341,** 1241214 (2013).
34. Antharam, V. C. *et al.* Intestinal Dysbiosis and Depletion of Butyrogenic Bacteria in Clostridium difficile Infection and Nosocomial Diarrhea. *J. Clin. Microbiol.* **51,** 2884–2892 (2013).
35. Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *PNAS* **104,** 13780–13785 (2007).
36. Shukla, R., Ghoshal, U., Dhole, T. N. & Ghoshal, U. C. Fecal Microbiota in Patients with Irritable Bowel Syndrome Compared with Healthy Controls Using Real-Time Polymerase Chain Reaction: An Evidence of Dysbiosis. *Dig. Dis. Sci.* (2015). doi:10.1007/s10620-015-3607-y
37. Palma, G. D. *et al.* Intestinal dysbiosis and reduced immunoglobulin-coated bacteria associated with coeliac disease in children. *BMC Microbiology* **10,** 63 (2010).
38. Yan, A. W. *et al.* Enteric dysbiosis associated with a mouse model of alcoholic liver disease. *Hepatology* **53,** 96–105 (2011).
39. Rea, M. C. *et al.* Effect of broad- and narrow-spectrum antimicrobials on Clostridium difficile and microbial diversity in a model of the distal colon. *PNAS* **108,** 4639–4644 (2011).
40. Kassam, Z., Lee, C. H., Yuan, Y. & Hunt, R. H. Fecal microbiota transplantation for Clostridium difficile infection: systematic review and meta-analysis. *Am. J. Gastroenterol.* **108,** 500–508 (2013).
41. Rea, M. C. *et al.* Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against Clostridium difficile. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 9352–9357 (2010).
42. Jarchum, I., Liu, M., Lipuma, L. & Pamer, E. G. Toll-like receptor 5 stimulation protects mice from acute Clostridium difficile colitis. *Infect. Immun.* **79,** 1498–1503 (2011).
43. Jarchum, I., Liu, M., Shi, C., Equinda, M. & Pamer, E. G. Critical role for MyD88-mediated neutrophil recruitment during Clostridium difficile colitis. *Infect. Immun.* **80,** 2989–2996 (2012).
44. Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502,** 96–99 (2013).
45. Buffie, C. G. *et al.* Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. *Nature* **517,** 205–208 (2015).
46. Wang, I. J. Examining the Full Effects of Landscape Heterogeneity on Spatial Genetic Variation: A Multiple Matrix Regression Approach for Quantifying Geographic and Ecological Isolation. *Evolution* **67,** 3403–3411 (2013).