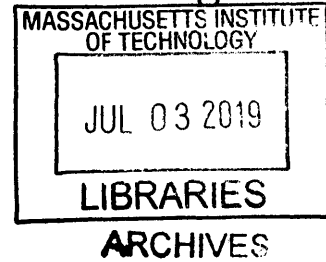# Mining the human microbiome for clinical insight

by

Claire Marie Noëlle Duvallet

B.S., Columbia University (2013)

Submitted to the Department of Biological Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

© Massachusetts Institute of Technology 2019. All rights reserved.

Author . . . . . . **Signature redacted** . . . .

Department of Biological Engineering
11 January 2019

Certified by . . . . . . **Signature redacted** . . . .

Eric J. Alm
Professor
Thesis Supervisor

Accepted by . . . . . . **Signature redacted** . . . .

Forest White
Chair of Graduate Program, Department of Biological Engineering

Mining the human microbiome for clinical insight

by

Claire Marie Noëlle Duvallet

Submitted to the Department of Biological Engineering
on 11 January 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Biological Engineering

Abstract

The human microbiome is essential for health and has been implicated in many diseases. DNA sequencing has enabled the detailed characterization of these human-associated microbial communities, leading to a rapid expansion in studies investigating the human microbiome. In this thesis, I describe multiple projects which overcome various data analysis challenges to extract useful clinical insights from microbiome data. In the first project, I present an analysis of lung, stomach, and oropharyngeal microbiomes. I leverage data collected from multiple sites per patient to identify aspiration-associated changes in the relationships between these communities, discovering new properties of the aerodigestive microbiome and suggesting new approaches for treatment. In the second project, I perform a meta-analysis of case-control gut microbiome datasets with standard data processing and analysis methods. I find consistent patterns characterizing disease-associated microbiome changes and a set of shared associations which could inform clinical treatment and therapeutic development approaches for different microbiome-mediated diseases. Enabled by this work, in the third project I contribute to the development of a method to correct for batch effects in case-control microbiome studies. In the fourth project, I describe a framework for rational donor selection in fecal microbiota transplant clinical trials in which knowledge derived from clinical and basic science research is used to inform which donor is selected for fecal transplants, increasing the likelihood of successful trials. Finally, I present preliminary results analyzing the microbiome and metabolome of residential sewage as a novel platform for community-level public health surveillance. Together, these projects demonstrate a variety of approaches to mine the human microbiome for clinically-relevant insights and suggests multiple avenues forward for translating findings from microbiome data analyses into clinical and public health impact.

Thesis Supervisor: Eric J. Alm
Title: Professor

Acknowledgments

I look back on my PhD with almost exclusive positivity: the road was rough at times, but graduate school was an overwhelmingly enriching and positive experience. I recognize that this is a rare sentiment to have upon graduation, and I credit my communities, role models, mentors, family, and friends for this seemingly impossible feat.

If I have thrived in grad school, it is in large part because of my advisor, Eric Alm. He waters whatever seeds his students decide to plant, and I am deeply grateful for him giving me the support I needed in my research and the space I needed to succeed in my non-scientific endeavors. Learning how to do science from Eric has given me the confidence to tackle hard problems, jump into creative solutions, and figure out the details later. Thanks, above all, for believing in me and reminding me of that often enough for it to stick.

I am also grateful to my committee, who was a constant source of support and never one of stress. Rafael Irizarry welcomed me into his group and gave his mentorship freely, even when my projects did not fall within his direct expertise. I've learned so much from his perspectives on statistics and data science, which I will keep beyond my PhD. I hope to always remain a friend of the lab. Deborah Hung provided a much-appreciated clinical and public health perspective to my work and career path. And finally thanks to my committee chair, Jim Collins, for providing realism and unequivocal support throughout my PhD.

Rachel Rosen was the clinical PI on the aerodigestive microbiome project, and I thank her for sharing her data and knowledge and, most importantly, her passion for her patients. Working with her – a real doctor! – was an incredible privilege, and provided me with important perspectives on the realities of doing actual translational research. Elizabeth Kujawinski was a mentor and colleague on the sewage project, and taught me basically everything I know about metabolomics. I thank her for her generosity and for infecting me with her passion for chemistry.

There are many postdocs and senior graduate students whom I credit for putting me onto a path to success. Ilana Brito was a mentor I didn't know I needed, and inspired me to hustle through this last year of my PhD to find my way to (and hopefully through) an untraditional career path. Sarah Spencer also provided key "unsolicited advice" early in my rotation and thesis proposal processes which was incredibly helpful. Her laughter is infectious and her legacy of generosity in the lab is one I aspire to. Thomas Gurry was my first introduction to computational work during my rotation, and gave me much more time and energy than was required of him. He provided me a solid computational foundation and was the first to introduce me to the possibility of a PhD with only computational work, which would become my passion. I worked closely with Mariana Matus on the Underworlds sewage project, which was key to helping me identify public health as a field where my passions and skills intersect. Learning from her entrepreneurial spirit convinced me that there are no limits to what can be done in graduate school. Scott Olesen was many things to me throughout my PhD: a role model, mentor, collaborator, and friend. He was a

crucially positive force in almost every aspect of my PhD: starting me on my path to student advocacy, sharing tips and tricks (and rants!) about computing, and being a model of a grad student to aspire to. There are few people in my life who have had an impact like Scott's on shaping who I am. Finally, I send infinite thanks to Sean Gibbons, my close collaborator on the meta-analysis and batch correction projects. Sean is such a joy to work with, so purely believes in and loves science, and re-inspired me through my existential crises when I stopped thinking that data meant anything at all. Beyond being just a mentor or role model, Sean is a true sponsor: volunteering me for important opportunities and always at the ready to boost my scientific career. If more scientists (and people!) were like Sean, the world would be a better place.

I am deeply grateful to the entire Alm lab for being a community of laughter, curiosity, and generosity. The past few years have been a joy and it is in large part because of the culture that is cultivated and sustained in the group. I especially could not imagine my PhD without Nathaniel Chu, who was an amazing office mate and provided endless patience and insights for our many discussions ranging across many topics – scientific and otherwise. Thanks also for the infinite guacamole and all the M&M's. Mathieu Groussin was also an excellent compatriot in Office 307 – thanks for the jokes, wisdom, and excellent haircuts. Finally, no one in the Alm lab has any chance at success without Shandrina Burns and I thank her for her flexibility, support, and patience. She truly works magic.

I was also happy to become a friend of the Rafalab, and I thank the entire group for welcoming me and for giving me access to a community where computationalists are the norm. Though they may not realize it, Stephanie Hicks and Keegan Korthauer are important role models of mine, as thoughtful and positive people who stick up for what's right and strong women who do computational work. Along with Patrick Kimes and Alejandro Reyes, I also thank them for humoring my questions about R and patiently teaching me new things in spite of my commitment to Python.

Beyond the work presented in these pages, I spent a significant portion of my PhD in student leadership roles in my department and at MIT. Learning conflict management skills as a BE REF positively impacted almost every aspect of my life. This impact wouldn't have happened without the people involved in my early REFS days: Libby Mahaffy, Scott Olesen, Georgia Lagoudas, Tú Nguyen, Deena Rennerfeldt, and Souparno Ghosh. I also thank my early diversity co-conspirators, Scott Olesen and Manu Kumar, for sharing in my rants and helping me channel my passions toward productive change. Amanda Facklam was also key to getting the BE Application Assistance Program up and running, and I thank her for her sustained support and enthusiasm for this exciting initiative. I was able to grow tremendously as a student leader because of our receptive departmental leadership, and I am very thankful to Doug Lauffenburger for being thoughtful, supportive, and responsive to our efforts. My current GSC diversity crew, Nasir Almasri, Josué Lopéz, Danielle Orozco Cosio, German Parada, Halston Lim, and Bianca Lepe, revitalized me in this work when I was closest to burning out. The future of diversity and equity at MIT is bright if they are leaders in it. Similarly, I am grateful for the brief time I overlapped with Jess Boles and Parker Vascik in GradSAGE, and am excited to watch them positively transform the culture of advising at MIT.

Finally, I thank my friends, family, and frisbee teammates for keeping me happy and healthy during my PhD. I cannot thank Carolyn Saund and Megan Armstrong enough for being right by my side these past years, for supporting my growth, and for growing alongside me as we figure out how to maneuver this world. Hold on tight to your passion, and thanks in advance for reminding me to do the same. When I think about strong support networks of badass women, Andee Wallace and Shelby Doyle also come to mind. Maneuvering grad school together enriched my experience tremendously, kept me focused on what matters, and was a lot of fun. Thanks also to Brett Geiger, the BE 2014 cohort, and the entire BE community for being great friends and colleagues. I am also thankful to Janyne Little, who keeps me adventuring and reminds me about the world outside my urban ivory towers. Jeremy Pivor moving back to Boston was pivotal for my happiness in this city: I am so lucky for his friendship and treasure his radiant soul. I've also been lucky to share living spaces with incredibly supportive and friendly roommates: Kunal Bhutani, a man kinder and more thoughtful than I will ever be, and the 9 Seattleites, including my dear childhood friend (and yet another badass woman inspiration) Kelly Moynihan, as well as Khoi Nguyen, Andee Wallace, Ashvin Bashyam, Griffin Clausen, Santiago Correa, Shawn Musgrave, and of course Nancy, Barb, Sonia, and Ruth.

To everyone I have ever thrown a frisbee with: thank you for keeping me physically and mentally healthy. I am especially grateful to have been a member of sMITe, MIT's women's ultimate frisbee team, during my entire PhD. I treasure having had the opportunity to learn about MIT's fascinating undergraduate culture and to witness the transformation of so many young women into confident players and leaders. Lisa, Bethany, Margalit, Julia, Abby, Oreo, Kimmy, Kelly, Helen, Michelle, Mary, Amy, Theresa, Sam, Ivana, Maddi, Cynthia, Serena and everyone else on sMITe: thanks for putting up with me, and for being my friend.

I also thank my family for supporting and encouraging me my entire life. Felix for encouraging me to adopt life-changing habits like Getting Things Done and meditation; Etienne, for putting us to shame by having a real job and providing important perspective which grounded my research on IBD; and Maman et Papa, for always supporting me, recognizing my passions better than I do, and putting up with me talking about poop over many many dinners.

Contents

Chapter 1

Introduction

1.1 The human microbiome in health and disease

The microbes that live in and on our bodies make up the human microbiome and are essential for health. Almost all human body sites are colonized by microbes, ranging from the gut, the largest human-associated microbial community, to the lungs, which were for many years considered sterile [1, 2]. These microbial communities perform essential functions for health, including fighting off and preventing infections, regulating host metabolism and interacting with the immune system, and metabolizing xenobiotics or other compounds which are indigestible by the host. Additionally, perturbations in human microbiomes have been implicated in many diseases, including inflammatory, metabolic, neurological, and respiratory conditions [2, 3, 4, 5].

The potential for microbiome-based therapies to improve human health and address a broad range of diseases has led to a recent expansion of research and clinical studies in this field. Much of the emerging research has been driven in part by the increasing accessibility of DNA sequencing technology, which can provide a detailed view of the bacteria in these communities without the need for time-consuming and difficult culturing experiments [6]. However, identifying clinically-relevant associations from microbiome studies remains challenging [7]. Microbiome datasets are high-dimensional, with hundreds to thousands of bacterial species measured in usually only tens to hundreds of patient samples. Microbial communities are also

highly variable across people, making it more difficult to identify individual bacterial biomarkers that can consistently distinguish health and disease across many different patients. Finally, microbiome datasets provide a window into associations but not causal relationships, and researchers must perform mechanistic studies or clinical trials to confirm the clinical relevance of any identified associations. In this thesis, I present unique analyses of microbiome data which overcome some of these challenges and which illustrate a variety of approaches for extracting useful clinical insights from mining microbiome data. Together, the following chapters aim to move findings from the individual microbiome data analyses beyond statistical significance and toward clinical meaningfulness.

1.2 Multi-site sampling to identify clinical associations in the aerodigestive microbiome

In Chapter 2, I leverage simultaneous sampling within patients to identify clinically-relevant aerodigestive microbiome characteristics that distinguish patients with swallowing dysfunction from those with normal swallow. The current gold standard to diagnose aspiration resulting from impaired swallow function involves imaging a patient as they ingest different quantities and consistencies of contrast material [8], but identifying lung microbial biomarkers could provide a useful non-radioactive alternative to this diagnosis. Additionally, patients with impaired swallow function are at higher risk for respiratory infections, but the extent to which the lung microbiome is perturbed by aspiration and thus potentially involved in mediating this risk is unknown [8, 9]. Finally, clinical interventions to treat respiratory symptoms in patients with impaired swallow focus on preventing material transfer from the stomach into the lungs, for example via anti-reflux medication or fundoplication surgery [10, 11]. However, the clinical utility of these interventions in pediatric populations is not fully established, and the extent to which the gastric vs. mouth microbiome mediates respiratory complications is not known.

In this study, I analyze a set of lung, stomach, and oropharyngeal microbiome samples collected from over 200 pediatric patients at Boston Children's Hospital. I first show that lung and stomach microbiomes are highly variable across people and driven primarily by person rather than body site, complicating the search for a reliable microbial biomarker of aspiration across patients. I overcome this challenge by leveraging the fact that we have multiple samples per patient, comparing instead the within-patient *relationships* between aerodigestive microbial communities in patients with vs. without aspiration. Using this approach, I show that aspiration shifts lung microbial communities toward the oropharyngeal but not the stomach microbiome, suggesting that the mouth is likely an important source of lung microbiome perturbations in these patients. Thus, approaches for treating aspiration-related respiratory symptoms should target microbial transfer from the mouth into the lungs in addition to focusing on the gastric-lung axis, as most current interventions do. This study also illustrates the power of multi-site within-patient sampling to overcome variability across people to identify clinically meaningful microbiome-based biomarkers.

1.3 Re-analyzing datasets to find consistent patterns of associations between the gut microbiome and disease

In Chapter 3, I perform a meta-analysis of 28 case-control gut microbiome studies across 10 diseases to synthesize findings across studies and identify generalizable associations. Although the human gut microbiome has been extensively studied in many diseases, there is little consensus on which disease associations are consistent across patient cohorts. In other fields like medicine or psychology, consensus is usually achieved through meta-analyses of published literature [12]. However, comparing published results across microbiome studies is not straightforward. The field lacks standard data processing and analysis methods, making many reported results impossible or inappropriate to compare directly between studies. For example, differ-

9

ent studies often use incompatible bioinformatics or statistical methods: they may compare bacteria at different taxonomic levels or use different bioinformatics workflows, and they may identify significant associations through univariate statistical tests or machine learning models, results from which can not be readily compared [13, 14, 15, 16]. Additionally, studies led by clinicians often ask very different questions of the data than studies led by microbial ecologists, and so the reported results are not necessarily representative of the comprehensive information contained in the full dataset [17, 18, 19]. This issue is especially relevant in this field, where non-invasive sample collection and open-source bioinformatics software suites have made microbiome research accessible to a broad range of scientists and clinicians [20].

In this chapter, I perform a meta-analysis of gut microbiome studies which overcomes many of these challenges by reprocessing and reanalyzing raw data with standard methods. Even though specific bacterial associations vary across studies of the same disease, I identify patterns of general microbiome shifts which are consistent and which each suggest different approaches for developing microbiome-based therapeutics. When looking across multiple diseases, I find a set of bacteria which are non-specifically associated with health and disease, perhaps forming a shared or core response to general health and disease status. These results highlight the importance of contextualizing results from individual studies within the existing body of work, and also hints at the possibility of developing broadly beneficial targeted probiotic and antibiotic therapies.

1.4 Enabling more powerful meta-analyses by developing a batch correction method for microbiome data

Motivated and enabled by the work presented in Chapter 3, I present a method to correct for batch effects in case-control microbiome studies in Chapter 4. Traditionally, meta-analyses compile published p-values and effect sizes and apply statistical

methods to determine their consistency, significance, and magnitude across studies in order to glean an overall understanding of true effects [12]. However, this is challenging to do in the microbiome research field for reasons described above and in Chapter 3. A more powerful way to synthesize findings across studies is to combine the raw data and re-perform analyses on this expanded sample size. However, analyzing raw data combined from multiple studies is not appropriate without first correcting for batch effects. In microbiome data, batch effects result from biological and technical variation between studies and although statistical methods to correct for these have been developed for other 'omics data types, few are applicable for microbiome data [21].

In this chapter, we describe a method to correct for batch effects in case-control microbiome studies. Briefly, the method converts the abundances of taxa in case samples to percentiles of their respective distributions in the control samples within each study, in essence considering control samples as the null. Assuming that control patients represent biologically similar groups, this allows for pooling of percentile-normalized data across studies. We show that this non-parametric method successfully removes batch effects in case-control studies, enables more powerful analyses on combined microbiome data, and reduces the number of false positives from considering studies individually. My main contributions to this joint work were processing the datasets used to develop and test the method and implementing the method into an open-source microbiome software suite, expanding its reach and accessibility [20].

1.5 Translating microbiome research through rationally designed fecal microbiota transplant clinical trials

In Chapter 5, I present a framework for rationally selecting donors in fecal microbiota transplant (FMT) clinical trials. In recurrent *Clostridium difficile* infection (rCDI), FMTs have proven themselves as a remarkably effective clinical treatment, generating

11

excitement about the clinical potential of FMTs in other diseases [22, 23]. Clinical trials in non-rCDI diseases can be used to identify promising conditions in which the microbiome may have a causal role to play. However, although there is little donor-dependent variability in FMT efficacy in rCDI, it is becoming apparent that in many other conditions, donor heterogeneity likely plays a role in patient response [24, 25]. Thus, as FMT trials expand to new indications, many trials may fail not because the indication was not amenable to improvement by FMT but rather because an ineffective donor was chosen. As a consequence, excitement and support for FMT research will be dampened, slowing progress toward finding clinical applications of microbiome-based therapies.

In this chapter, I present an approach for selecting donors in FMT trials to increase the likelihood that these trials will succeed. In this proposed framework, a clinician applies her existing knowledge and hypotheses for how her indication of interest is being mediated by the microbiome in order to drive the donor selection process. I present four types of disease models to describe underlying processes mediating microbiome-related diseases: acute dysbiosis, mediation by individual taxa, mediation by overall community function, and complex host-microbiome interactions. I suggest associated donor selection strategies for each of these disease models, and provide case studies illustrating the process of rational donor selection in practice. Finally, I perform a power simulation which finds that most FMT trials are unlikely to be powered enough to identify the key donor bacteria mediating patient responses to FMT. Thus, to successfully make discoveries from completed trials, clinicians should perform hypothesis-driven retrospective analyses and plan for these during the clinical trial design process, for example by collecting paired donor and patient and/or longitudinal samples. This framework encourages clinicians to leverage their clinical experience, existing microbiome research and published datasets, and the increasing availability of screened donor stools to more efficiently translate microbiome research into clinical impact. It also suggests approaches for improving retrospective analyses of the microbiome data generated during these clinical trials.

1.6 Mining the microbiome and metabolome of residential sewage for community-level public health surveillance

In Chapter 6, I present preliminary analyses of untargeted metabolomics and microbiome data from residential sewage, identifying human health and activity markers with potential public health relevance. Wastewater epidemiology has long been proposed for many applications in population health, but its implementation has so far been limited to polio surveillance [26] and illicit drug consumption monitoring [27, 28]. Standard wastewater epidemiology methods sample sewage at the wastewater treatment plant and perform targeted analyses to measure individual metabolites or viruses of interest. Sampling at the treatment plant can capture information about large urban populations, but does not provide spatially resolved information about individual communities within the city. Targeted analyses can confidently identify and quantify specific molecules, but cannot discover new biomarkers. Because a major goal of public health is to reduce disparities between populations by identifying and tailoring interventions to high-risk groups [29, 30, 31], limited geographical resolution and small sets of measurable biomarkers makes standard wastewater epidemiology methods inappropriate for many public health applications. Despite these limitations, adapting wastewater epidemiology and implementing it within communities could be useful for evaluating the impact of public health policies or interventions. For example, rather than waiting years to see how a citywide sugar tax affects obesity rates, biomarkers of human sugar consumption could be measured before and after the policy roll-out to directly and quickly evaluate its impact on sugar consumption in multiple neighborhoods.

In this chapter, we present a platform to implement wastewater epidemiology within cities that mines the metabolome and microbiome of residential sewage to find biomarkers with potential public health relevance. This project is the culmination of a large collaboration between computational biologists, urban designers,

engineers, city workers, and public health officials, and is presented in its entirety in Chapter 6. My contribution to this work was identifying and confirming the human biomarkers from the untargeted metabolomics data. By comparing the masses and fragmentation spectra of our metabolite features with databases and published work, I identified over 20 glucuronide compounds. These molecules are direct markers of human excretion and were abundant in our residential sewage samples but rarely detected at our downstream sampling site. In collaboration with a mass spectrometry specialist, I also confirmed multiple urinary and fecal metabolites and showed that their abundances reflected expected human activity patterns over the course of a 24-hour sampling. Further putative matches between our metabolomics features and a large database of human metabolites [32] suggests that this work could be extended to identify many different types of biomarkers reflecting human health, diet, and lifestyle. Together, this work shows that residential sewage contains more human-derived chemical and biological information than downstream samples, and suggests that upstream wastewater epidemiology could be developed into a new tool to measure community-level health.

Bibliography

[1] Ron Sender, Shai Fuchs, and Ron Milo. Are we really vastly outnumbered? revisiting the ratio of bacterial to host cells in humans. *Cell*, 164(3):337–340, 2016.

[2] James M. Beck, Vincent B. Young, and Gary B. Huffnagle. The microbiome of the lung. *Translational Research*, 160(4):258–266, oct 2012.

[3] Eliseo Papa, Michael Docktor, Christopher Smillie, Sarah Weber, Sarah P. Preheim, Dirk Gevers, Georgia Giannoukos, Dawn Ciulla, Diana Tabbaa, Jay Ingram, David B. Schauer, Doyle V. Ward, Joshua R. Korzenik, Ramnik J. Xavier, Athos Bousvaros, and Eric J. Alm. Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease. *PLoS ONE*, 7(6):e39242, jun 2012.

[4] V. K. Ridaura, J. J. Faith, F. E. Rey, J. Cheng, A. E. Duncan, A. L. Kau, N. W. Griffin, V. Lombard, B. Henrissat, J. R. Bain, M. J. Muehlbauer, O. Ilkayeva, C. F. Semenkovich, K. Funai, D. K. Hayashi, B. J. Lyle, M. C. Martini, L. K. Ursell, J. C. Clemente, W. Van Treuren, W. A. Walters, R. Knight, C. B. Newgard, A. C. Heath, and J. I. Gordon. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science*, 341(6150):1241214–1241214, sep 2013.

[5] Elaine Y. Hsiao, Sara W. McBride, Sophia Hsien, Gil Sharon, Embriette R. Hyde, Tyler McCue, Julian A. Codelli, Janet Chow, Sarah E. Reisman, Joseph F. Petrosino, Paul H. Patterson, and Sarkis K. Mazmanian. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell*, 155(7):1451–1463, dec 2013.

[6] The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402):207–214, jun 2012.

[7] Dan Knights, Laura Wegener Parfrey, Jesse Zaneveld, Catherine Lozupone, and Rob Knight. Human-associated microbial signatures: Examining their predictive value. *Cell Host & Microbe*, 10(4):292–296, oct 2011.

[8] Ian J. Cook and Peter J. Kahrilas. AGA technical review on management of oropharyngeal dysphagia. *Gastroenterology*, 116(2):455–478, feb 1999.

[9] J. Thomson, M. Hall, L. Ambroggio, B. Stone, R. Srivastava, S. S. Shah, and J. G. Berry. Aspiration and non-aspiration pneumonia in hospitalized children with neurologic impairment. *Pediatrics*, 137(2):e20151612–e20151612, jan 2016.

[10] A. B. Goldin, R. Sawin, K. D. Seidel, and D. R. Flum. Do antireflux operations decrease the rate of reflux-related hospitalizations in children? *Pediatrics*, 118(6):2326–2333, dec 2006.

[11] Steven L. Lee, Hooman Shabatian, Jin-Wen Hsu, Harry Applebaum, and Philip I. Haigh. Hospital admissions for respiratory symptoms and failure to thrive before and after nissen fundoplication. *Journal of Pediatric Surgery*, 43(1):59–65, jan 2008.

[12] Gene V. Glass. Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10):3–8, nov 1976.

[13] Pallavi Singh, Tracy K. Teal, Terence L. Marsh, James M. Tiedje, Rebekah Mosci, Katherine Jernigan, Angela Zell, Duane W. Newton, Hossein Salimnia, Paul Lephart, Daniel Sundin, Walid Khalife, Robert A. Britton, James T. Rudrik, and Shannon D. Manning. Intestinal microbial communities associated with acute enteric infections and disease recovery. *Microbiome*, 3(1), sep 2015.

[14] Nielson T. Baxter, Mack T. Ruffin, Mary A. M. Rogers, and Patrick D. Schloss. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Medicine*, 8(1), apr 2016.

[15] G. Zeller, J. Tap, A. Y. Voigt, S. Sunagawa, J. R. Kultima, P. I. Costea, A. Amiot, J. Bohm, F. Brunetti, N. Habermann, R. Hercog, M. Koch, A. Luciani, D. R. Mende, M. A. Schneider, P. Schrotz-King, C. Tournigand, J. Tran Van Nhieu, T. Yamada, J. Zimmermann, V. Benes, M. Kloor, C. M. Ulrich, M. von Knebel Doeberitz, I. Sobhani, and P. Bork. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Molecular Systems Biology*, 10(11):766–766, nov 2014.

[16] Margaret L. Zupancic, Brandi L. Cantarel, Zhenqiu Liu, Elliott F. Drabek, Kathleen A. Ryan, Shana Cirimotich, Cheron Jones, Rob Knight, William A. Walters, Daniel Knights, Emmanuel F. Mongodin, Richard B. Horenstein, Braxton D. Mitchell, Nanette Steinle, Soren Snitker, Alan R. Shuldiner, and Claire M. Fraser. Analysis of the gut microbiota in the old order amish and its relation to the metabolic syndrome. *PLoS ONE*, 7(8):e43052, aug 2012.

[17] Joshua S. Son, Ling J. Zheng, Leahana M. Rowehl, Xinyu Tian, Yuanhao Zhang, Wei Zhu, Leighann Litcher-Kelly, Kenneth D. Gadow, Grace Gathungu, Charles E. Robertson, Diana Ir, Daniel N. Frank, and Ellen Li. Comparison of fecal microbiota in children with autism spectrum disorders and neurotypical siblings in the simons simplex collection. *PLOS ONE*, 10(10):e0137725, oct 2015.

[18] Jose U Scher, Andrew Sczesnak, Randy S Longman, Nicola Segata, Carles Ubeda, Craig Bielski, Tim Rostron, Vincenzo Cerundolo, Eric G Pamer, Steven B Abramson, Curtis Huttenhower, and Dan R Littman. Expansion of intestinal prevotella copri correlates with enhanced susceptibility to arthritis. *eLife*, 2, nov 2013.

[19] Vincent Wai-Sun Wong, Chi-Hang Tse, Tommy Tsan-Yuk Lam, Grace Lai-Hung Wong, Angel Mei-Ling Chim, Winnie Chiu-Wing Chu, David Ka-Wai Yeung, Patrick Tik-Wan Law, Hoi-Shan Kwan, Jun Yu, Joseph Jao-Yiu Sung, and Henry Lik-Yuen Chan. Molecular characterization of the fecal microbiota in patients with nonalcoholic steatohepatitis – a longitudinal study. *PLoS ONE*, 8(4):e62885, apr 2013.

[20] Evan Bolyen, Jai Ram Rideout, Matthew R Dillon, Nicholas A Bokulich, and et al. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. dec 2018.

[21] Sean M. Gibbons, Claire Duvallet, and Eric J. Alm. Correcting for batch effects in case-control microbiome studies. *PLOS Computational Biology*, 14(4):e1006102, apr 2018.

[22] M. N. Quraishi, M. Widlak, N. Bhala, D. Moore, M. Price, N. Sharma, and T. H. Iqbal. Systematic review with meta-analysis: the efficacy of faecal microbiota transplantation for the treatment of recurrent and refractory clostridium difficile infection. *Alimentary Pharmacology & Therapeutics*, 46(5):479–493, jul 2017.

[23] AÃŕda Bafeta, Amelie Yavchitz, Carolina Riveros, Rui Batista, and Philippe Ravaud. Methods and reporting studies assessing fecal microbiota transplantation. *Annals of Internal Medicine*, 167(1):34, may 2017.

[24] Paul Moayyedi, Michael G. Surette, Peter T. Kim, Josie Libertucci, Melanie Wolfe, Catherine Onischi, David Armstrong, John K. Marshall, Zain Kassam, Walter Reinisch, and Christine H. Lee. Fecal microbiota transplantation induces remission in patients with active ulcerative colitis in a randomized controlled trial. *Gastroenterology*, 149(1):102–109.e6, jul 2015.

[25] Scott W. Olesen, McKenzie M. Leier, Eric J. Alm, and Stacy A. Kahn. Searching for superstool: maximizing the therapeutic potential of FMT. *Nature Reviews Gastroenterology & Hepatology*, 15(7):387–388, apr 2018.

[26] Y Manor, L M Shulma, E Kaliner, M Hindiyeh, D Ram, D Sofer, J Moran-Gilad, B Lev, I Grotto, R Gamzu, and E Mendelson. Intensified environmental surveillance supporting the response to wild poliovirus type 1 silent circulation in israel, 2013. *Eurosurveillance*, 19, 2014.

[27] Bikram Subedi and Kurunthachalam Kannan. Mass loading and removal of select illicit drugs in two wastewater treatment plants in new york state and estimation

of illicit drug usage in communities through wastewater analysis. *Environmental Science & Technology*, 48(12):6661–6670, jun 2014.

[28] Christoph Ort, Alexander L. N. van Nuijs, Jean-Daniel Berset, Lubertus Bijlsma, Sara Castiglioni, Adrian Covaci, Pim de Voogt, Erik Emke, Despo Fatta-Kassinos, Paul Griffiths, Félix Hernández, Iria González-Mariño, Roman Grabic, Barbara Kasprzyk-Hordern, Nicola Mastroianni, Axel Meierjohann, Thomas Nefau, Marcus Östman, Yolanda Pico, Ines Racamonde, Malcolm Reid, Jaroslav Slobodnik, Senka Terzic, Nikolaos Thomaidis, and Kevin V. Thomas. Spatial differences and temporal changes in illicit drug use in europe quantified by wastewater analysis. *Addiction*, 109(8):1338–1352, may 2014.

[29] Ramya Ramaswami, Ronald Bayer, and Sandro Galea. Precision medicine from a public health perspective. *Annual Review of Public Health*, 39(1):153–168, apr 2018.

[30] Tarun Stephen Weeramanthri, Hugh J. S. Dawkins, Gareth Baynam, Matthew Bellgard, Ori Gudes, and James Bernard Semmens. Editorial: Precision public health. *Frontiers in Public Health*, 6, apr 2018.

[31] Muin J. Khoury and Sandro Galea. Will precision medicine improve population health? *JAMA*, 316(13):1357, oct 2016.

[32] David S. Wishart, Timothy Jewison, An Chi Guo, Michael Wilson, Craig Knox, Yifeng Liu, Yannick Djoumbou, Rupasri Mandal, Farid Aziat, Edison Dong, Souhaila Bouatra, Igor Sinelnikov, David Arndt, Jianguo Xia, Philip Liu, Faizath Yallou, Trent Bjorndahl, Rolando Perez-Pineiro, Roman Eisner, Felicity Allen, Vanessa Neveu, Russ Greiner, and Augustin Scalbert. HMDB 3.0—the human metabolome database in 2013. *Nucleic Acids Research*, 41(D1):D801–D807, nov 2012.

Chapter 2

Multi-site sampling reveals altered relationships in aerodigestive microbiomes of children with swallowing dysfunction

Claire Duvallet, Kara Larson, Scott Snapper, Sonia Iosim, Ann Lee, Katherine Freer, Kara May, Eric J Alm, and Rachel Rosen.

The contents of this chapter have been submitted for publication as: "Aerodigestive sampling reveals altered microbial exchange between lung, oropharyngeal, and gastric microbiomes in children with impaired swallow function," in review as of December 2018.

The tables and figures are at the end of the Chapter. The Supplementary Information is in Appendix A.

Abstract

Background

Children with oropharyngeal dysphagia have impaired airway protection mechanisms and are at higher risk for pneumonia and other pulmonary complications. Aspiration of gastric contents is often implicated as a cause for these pulmonary complications, despite being supported by little evidence. The goal of this study is to determine the relative contribution of oropharyngeal and gastric microbial communities to perturbations in the lung microbiome of children with and without oropharyngeal dysphagia and aspiration.

Methods

We conducted a prospective cohort study of 222 patients consecutively recruited from a tertiary aerodigestive center undergoing simultaneous esophagogastroduodenoscopy and flexible bronchoscopy. Bronchoalveolar lavage, gastric and oropharyngeal samples were collected and 16S sequencing was performed. A subset of patients also underwent video fluoroscopic swallow studies to assess swallow function and were categorized as aspiration/no aspiration. Microbial communities across the aerodigestive tract were compared in patients with and without aspiration by calculating within-patient beta diversities and quantifying microbial exchange across sites.

Results

Within all patients, lung, oropharyngeal and gastric microbiomes overlap. The degree of similarity is the lowest between the oropharynx and lungs (median Jensen-Shannon distance (JSD) = 0.90), and as high between the stomach and lungs as between the oropharynx and stomach (median JSD = 0.55 and 0.56, respectively; p = 0.6). Unlike the oropharyngeal microbiome, lung and gastric communities are highly variable across people and driven primarily by person rather than body site. In patients with aspiration, the lung microbiome more closely resembles oropharyngeal rather than

gastric communities and there is greater prevalence of microbial exchange between the lung and oropharynx than between gastric and lung sites ($p = 0.04$ and $3x10^{-5}$, respectively).

Conclusions

The gastric and lung microbiomes display significant overlap in patients with intact airway protective mechanisms while the lung and oropharynx remain distinct. In patients with impaired swallow function and aspiration, the lung microbiome shifts towards oropharyngeal rather than gastric communities. This finding may explain why antireflux surgeries fail to show benefit in pediatric pulmonary outcomes.

2.1 Introduction

The economic and social impact of oropharyngeal dysfunction and aspiration is well known in the adult stroke population; adults with oropharyngeal dysfunction are at greater risk of pneumonia than those without [1]. Little is known about aspiration-related lung disease in children, though recent studies suggest that up to 10% of all pneumonia hospitalizations in pediatrics are related to aspiration [2]. Clinicians often assume these pneumonias result from the aspiration of refluxed gastric contents and frequently treat these children with antireflux surgery, fundoplication. Despite this common surgical practice, there are no pediatric studies which conclusively show improved pulmonary outcomes after fundoplication, suggesting that the respiratory symptoms seen in aspirating patients may not be related to aspiration of gastric contents [3, 4, 5, 6, 7]. An alternative hypothesis is that aspiration-related respiratory symptoms may result from aspirated oropharyngeal contents. To test this hypothesis, we determined the microbial signatures of the lungs, stomach, and oropharynx in children with and without oropharyngeal dysphagia (i.e. with and without impaired airway protective mechanisms) to determine the relative contributions of the oropharyngeal and gastric microbiomes to the lung microbiome.

Previous studies have shown that the mouth, upper respiratory tract, and lung microbiota contain similar microbes, and that upstream oral communities seed downstream sites (e.g. lungs and stomach) [8, 9, 10]. However, there is little consensus on whether there exists a distinct or "core" lung microbiome that is consistent across people [9, 11, 12, 13]. Most studies, however, agree that the lung microbial communities share taxa with the oral microbiome, but that there are some bacteria present in lung communities whose abundances cannot be traced solely to the mouth [8, 9, 14, 12].

While the importance of oropharyngeal flora in seeding the lungs has been heavily studied in ICU settings [15, 16, 17], the role of oropharyngeal-lung flora exchange in otherwise heathy children with isolated swallowing dysfunction is unknown. Furthermore, studies investigating the relationships between microbial communities across the aerodigestive tract have not examined how microbes exchange between the stom-

ach and lungs, and how this exchange relates to clinical factors such as aspiration and gastroesophageal reflux.

If the lung microbiome of aspirating patients exhibits more exchange with the oropharynx than the stomach, this could provide evidence for why anti-reflux surgery is not helpful in patients with aspiration-related respiratory symptoms. Furthermore, a shift in the lung microbial communities toward an oropharyngeal population could not only result in overt pneumonia but may also have more subtle, pro-inflammatory effects [18]. Finally, if there is a unique aerodigestive microbial signature in aspirating patients, microbial profiling may be helpful as a diagnostic tool for oropharyngeal dysphagia.

2.2 Methods

2.2.1 Patient cohort and sample collection

We conducted a prospective cross sectional cohort study of children ages 1–18 undergoing bronchoscopy and esophagogastroduodenoscopy (EGD) for the evaluation of chronic cough. Patients with gastrostomy or nasogastric tubes, a history of gastrointestinal surgery, or antibiotics within 4 weeks of sample acquisition were excluded. The study was approved by the Boston Children's Hospital Institutional Review Board and informed consent was obtained from all patients/parents.

We first performed brushing of the posterior tongue to obtain oropharyngeal samples, placing the brush in TE buffer at -80C. Second, the bronchoscopy and bronchoalveolar lavage (BAL) was performed through an endotracheal tube in distal airways of the right middle lung or the most visually inflamed lung. Finally, gastric sampling was performed during the EGD. The endoscope was advanced, without suctioning, immediately into the stomach where the gastric fluid was suctioned into a sterile leukitrap. A minimum of 1 cc of gastric and lung fluid were collected and transferred to -80C. Each patient had a triad of samples collected: oropharynx, gastric fluid, and BAL (Tables 2.1 and 2.2) [10].

23

2.2.2 Multichannel intraluminal impedance with pH (pH-MII)

A subset of patients had pH-MII testing at the discretion of the patient's primary gastroenterologist. Acid reflux episodes were defined as episodes detected by the impedance (MII) sensors with associated drop in pH to < 4; non-acid episodes did not have the associated drop. The percentage of time that reflux was in the proximal/distal esophagus was calculated by dividing the sum of the bolus clearance times in the proximal/distal esophagus by the total study duration. The percentage of full column reflux events was defined as the percentage of the total reflux events that reached the proximal two impedance sensors (i.e., the proximal most impedance channel) [19].

2.2.3 Oropharyngeal dysphagia assessment

A subset of the patients included in this study had a videofluoroscopic swallow study (VFSS) to asses swallow function and were divided into two groups (normal swallow function and aspiration/penetration). Because patients with penetration on VFSS have similar pulmonary symptoms and respond similarly to thickening as patients that aspirate, we included patients with aspiration and penetration in one group.

2.2.4 Sample processing and sequencing

Oropharyngeal swabs, BAL, and gastric fluid samples suspended in Tris-Saline buffer were centrifuged for 3 minutes at 10,000 rcf prior to DNA isolation. DNA was extracted from the sample pellet with the Qiagen DNeasy PowerSoil Kit as described by the manufacturer, with the following modifications: protein precipitation in one step using 100 μL of each C2 and C3 solutions, and column centrifugation at 10,000 rcf for 10 minutes. Sequencing was performed in two batches at the Broad Institute. Patients with multiple samples had all of their respective samples sequenced in the same batch.

2.2.5 Microbiome data processing and analysis

Paired end reads were merged using USEARCH -fastq_mergepairs and truncated to 200 bp. Reads with more than 2 expected errors were discarded. Operational taxonomic units (OTUs) were clustered at 99% similarity and assigned taxonomy using the RDP classifier (c = 0.5) [20]. All quality filtering and OTU calling steps were performed with an in-house pipeline (https://github.com/thomasgurry/amplicon_sequencing_pipeline).

Beta diversity was calculated with the Jensen-Shannon distance (JSD). Only samples which were sequenced in the same batch were considered in cross-patient comparisons. Differences in overall community structure across sites was assessed using the PERMANOVA test as implemented in scikit-bio v 0.4.2 (skbio.stats.distance.permanova).

To define exchanged OTUs, we used data from patients with all three sites sequenced. For each OTU, we calculated the Spearman partial correlation ($\frac{r_{xy} - r_{xz}r_{zy}}{\sqrt{(1-r_{xz}^2)(1-r_{zy}^2)}}$) between its non-zero abundances in two sites, partialled on the third site (Scipy v 0.19.0 stats.spearmanr). P-values for each OTU were calculated as the percentage of null correlations larger than the observed correlation after shuffling abundances 2000 times. Only OTUs present in two sites in at least 10 patients were considered. OTUs with FDR-corrected q-value < 0.1 were defined as "exchanged" (sandbox.stats.multicomp.multipletests with method='fdr_bh'). To determine the statistical significance of the number of exchanged OTUs, we shuffled the patient IDs for each OTU in each site and re-defined "null" exchanged OTUs as described above.

We used five-fold cross-validation and Random Forest classifiers (scikit-learn v 0.18.1 ensemble.RandomForestClassifier with n_estimators=1000) for all supervised machine learning analyses. Areas under the ROC curve (AUCs) were calculated based on the predictions on each fold's test set (mean values across folds is reported) and Fisher p-values were calculated from all test set predictions. The aspiration/non-aspiration classifiers varied different train/test splits, so we report the mean results

across 100 repetitions.

2.2.6 Availability of data and materials

Code to reproduce the analyses presented here are available at www.github.com/cduvallet/aspiration-analysis-public. The 16S sequencing data used in this study will be made available in the SRA repository at accession number SRP141148 upon publication and clinical metadata are available upon request from the corresponding author.

2.3 Results

Two hundred and twenty two patients were included in the analysis (Tables 2.1 and 2.2). The mean age of the patients was 7.1 ± 5.4 years. One hundred out of 222 patients were taking proton pump inhibitors at the time of sampling. One hundred and four patients had a videoflouroscopic swallow study of which 47 (45%) had evidence of aspiration or penetration and 57 (55%) had normal swallow function. Thirty one patients had pH-MII testing for gastroesophageal reflux at the time of sample collection.

2.3.1 Aerodigestive microbiome across people

At the genus level, pediatric aerodigestive communities share many predominant members, including *Streptococcus*, *Prevotella*, *Haemophilus*, *Veillonella*, and *Neisseria* (Figure 2-1). Despite genus-level similarities, OTU-level aerodigestive communities are distinct and highly variable across people. The overall community composition was significantly different between sites (PERMANOVA on JSD, $p < 0.001$, Figure 2-2B). Furthermore, lung communities were very different across people (median lung-lung JSD = 0.88) while oropharyngeal communities tended to be more similar (median oropharyngeal-oropharyngeal JSD = 0.59, Figure 2-2A).

2.3.2 Aerodigestive microbiome within people

We compared aerodigestive communities within patients who had multiple sites sequenced (Table 2.2, Figure 2-3). Oropharyngeal and gastric fluid communities are similar within patients (median JSD = 0.56), reflecting that the mouth seeds the gastric microbiome [8, 9]. The majority of patients had very different lung and oropharyngeal communities (median JSD = 0.90), and these differences were significantly higher than either the lung-gastric fluid or gastric fluid-oropharyngeal beta diversities ($p < 1 \times 10^{-8}$, Figure 2-3A). Surprisingly, lung and stomach communities were as similar to each other as stomach and oropharyngeal communities (median JSD = 0.55 versus median JSD = 0.56, respectively, p = 0.6).

To identify specific microbes exchanging between sites, we reasoned that an actively exchanging microbe's abundances in two sites should be correlated across patients (Supplementary Figure A-1 and Methods). We identified 12 OTUs exchanged between lung and oropharyngeal, 74 between gastric fluid and lung, and 118 between oropharyngeal and gastric fluid communities. These results were statistically significant: we found a maximum of 2 exchanged OTUs between sites in our null analysis. The low number of directly exchanged OTUs between the oropharynx and lungs supports the finding that these sites are more distinct than others in the aerodigestive tract. The lungs and stomach exchange fewer OTUs than the oropharynx and stomach even though they have comparable intra-patient similarities, suggesting that factors other than specific bacterial exchange contributes to the similarity between lungs and stomachs within patients.

Random Forest classifiers trained to distinguish between sites (ensuring that samples from the same patient were in the same train/test set) were able to identify a generalizable oropharyngeal microbial signature that distinguishes the oropharynx from other sites across people (AUC = 0.95 for both gastric fluid and lung comparisons, Figure 2-3B). Surprisingly, when we compared within-patient similarities across sites to across-patient similarites for the same sites, we found that lung and stomach communities within patients were more similar than lungs across patients

27

and than stomachs across patients (Table 2.3, $p < 1 \times 10^{-8}$). Thus, while there exists a "core" oropharyngeal microbiome across people, lung and gastric communities are more variable and driven primarily by the person rather than body site. These results challenge the prevailing hypothesis that human-associated microbial communities are primarily driven by body habitat and instead suggest that patient-specific relationships may be equally, if not more, important in determining community structure in the aerodigestive microbiome [21, 22, 23].

2.3.3 Aspiration modulates the relationship between lung and oropharyngeal microbiomes but not the lung and stomach

Next, we investigated the impact of oropharyngeal dysphagia and aspiration on the relationships between aerodigestive microbiomes. Aspirators had significantly more similar lung and oropharyngeal communities than non-aspirators (Figure 2-4A, $p = 0.04$) and were much more likely to have the pre-defined oropharyngeal-lung microbes in both their oropharynx and lungs than non-aspirators ($p = 2 \times 10^{-5}$) (Figure 2-4B). Lung-oropharynx exchanged OTUs co-occurred in a median of 42% of aspirators' lung and oropharyngeal communities but only 20% of non-aspirators'. Aspirators were not more likely to have stomach-lung microbes present in both the lungs and gastric fluid than non-aspirators (Figure 2-4B, $p > 0.5$), and lung and gastric communities of aspirating patients were not necessarily more similar to each other than those of non-aspirating patients (Figure 2-4A, $p = 0.5$).

To identify potential microbial biomarkers of aspiration, we looked at the exchanged OTUs which were most frequently present in the lung and oropharyngeal communities of aspirators relative to non-aspirators. In the oropharyngeal-lung exchanged OTUs, these were an unknown OTU in the *Flavobacteriaceae* family, OTUs in the *Fusobacterium*, *Rothia*, *Veillonella* genera, and an unknown OTU in the *Prevotellaceae* family, among others (Table 2.4, gastric-lung OTUs in Supplementary Table A-1).

We used Random Forest classifiers trained on the presence of exchanged OTUs in different sites to test their potential as biomarkers. The concordant presence or absence of exchanged OTUs in the two sites improved classifiers based on the oropharyngeal-lung OTUs but not the ones based on the lung-gastric OTUs, relative to classifiers based on the presence of the exchanged OTUs in either site alone (Table 2.5). Using Random Forest classifiers trained on the entire microbiomes, we found that combining the oropharynx and lung communities resulted in a better classifier than either community alone (Table 2.6). Surprisingly, the classifiers trained on oropharyngeal and gastric communities performed well, despite our expectation that aspiration-induced changes in the microbiome would manifest in the lungs rather than the oropharynx or stomach. We confirmed that the patients' aspiration status was not confounded with proton pump inhibitor usage (Fisher exact p-value > 0.2), but there may be other co-morbidities or unmeasured confounders that could be driving the differences detected in these communities. However, taken together, these results suggest that identifying a biomarker for aspiration based on bacteria in both the lungs and oropharynx may be possible, and that these two sites together contain more information about a patient's aspiration status than either site alone.

2.3.4 Reflux may impact the relationship between lung and stomach microbiomes

Reflux profiles for the 31 patients are shown in Table 2.7. The percent of full column, distal, and proximal reflux were slightly negatively correlated with gastric-lung JSD, indicating that patients with more frequent reflux may have more similar gastric and lung microbial communities (Figure 2-5). However, the large range of gastric-lung JSDs across all patients and relatively weak correlation suggests that other non-reflux factors likely contribute more to the similarities between gastric and lung communities that are observed across all people.

2.4 Discussion

In this study, we characterized the relationships between the oropharyngeal, lung, and gastric microbiomes in a large pediatric cohort with and without swallowing dysfunction. Leveraging our simultaneous sampling of multiple sites per patient, we find that there exists a "core" oropharyngeal microbiome across patients, that the lung and gastric communities are highly variable across patients and driven primarily by patient rather than body site, and that within patients the lung and oropharyngeal communities remain most distinct. We show for the first time that in patients with impaired swallowing, the lung microbiome shifts toward oropharyngeal flora rather than gastric flora. Our results also suggest that identifying biomarkers for aspiration based on the presence of certain bacteria in both the lungs and oropharynx may ultimately be possible.

There are several limitations to our study. First, because it is unethical to perform bronchoscopies on healthy children, our patients in this study had respiratory symptoms. However, we believe that our patient population represents patients typically seen in aerodigestive centers and that understanding the degree of microbial exchange is most clinically relevant in patients with symptoms. The microbial populations we found in this study are similar to those of previously published studies of both healthy and symptomatic adults which reinforces the validity of our results [8, 9, 13, 14]. Second, the number of patients undergoing pH-MII testing was relatively small which limits our conclusions about the impact of gastroesophageal reflux on the lung. However, our study raises enough concerns about the significance of oropharyngeal-lung exchange in children with impaired swallowing that gastroesophageal reflux should not be considered as the primary source of microbial exchange causing pulmonary symptoms. Third, the diagnosis of oropharyngeal dysphagia in this study was based on VFSS. While this only categorizes patients based on a "one-point-in-time" study, it is the gold standard test to diagnose oropharyngeal dysphagia in children and therefore we feel it is appropriate for use in this study.

Despite these limitations, our findings have broad clinical implications for the

understanding and treatment of oropharyngeal dysphagia with resultant aspiration. Our clinical finding that the lung microbiome in children with aspiration shifts toward the oropharynx rather than the stomach highlights the importance of understanding the primary driver of microbial exchange so that therapies can be tailored accordingly. For example, if the mechanism of lung symptoms and disease in aspirating children results from a microbial shift towards oropharyngeal flora, anti-reflux surgery will be of no benefit to preventing oropharyngeal-lung exchange. Instead, therapies may need to be tailored to focused on changing oropharyngeal flora or salivary properties.

While there are no existing pediatric microbiome studies of the aerodigestive microbiome in patients with dysphagia, there is evidence that children with oropharyngeal dysphagia are predisposed to pneumonia and that this could be due to increased aspiration of microbes from the oral microbiome. In a study of 382 children undergoing VFSS, evidence of aspiration predicted pneumonia risk, though the causative organisms for these pneumonias were not known [24]. In cohort of elderly aspirating patients, oral colonization by respiratory pathogens was associated with increased risk of pneumonia, highlighting the potential importance of oral flora in influencing the lung outcomes [25]. Finally, a previous study of healthy adults found that individuals with oropharyngeal bacteria in their lungs had increased evidence of inflammatory metabolomic signals, suggesting that even a change of lung flora to commensal oropharyngeal bacteria can trigger inflammation even in healthy patients [18]. Our results add to these findings and suggest that changes in the lung microbiome towards oropharyngeal flora merit additional study to determine if these shifts result in increased morbidity or worse clinical outcomes, including the development of pneumonia.

From a microbial perspective, we identified bacterial families and genera that are more commonly exchanged between the oropharynx and lungs of children that aspirate than of children with intact swallowing mechanism. While there are no other 16S sequencing studies determining aspiration pneumonia risk in children, there is evidence from the adult literature that similar bacteria are involved in aspiration pneumonia risk. For example, oropharyngeal *Streptococci* were found to be more

abundant in the lungs of adults with pneumonia and aspiration risk factors than without aspiration risk [26]. In a study of 173 adults in long term care facilities, patients with oropharyngeal *Prevotella* and *Veillonella* had increased risk of death from pneumonia compared to patients who had oropharyngeal *Neisseria* and *Fusobacterium* [27]. Our study is a critical first step toward identifying bacteria present in the oropharynx and lungs of aspirating children that may result in higher risk for pneumonias, with additional studies needed to determine their impact on pediatric outcomes.

In summary, our findings suggest that interventions to reduce aspiration-related respiratory complications due to increased microbial exchange should target aspiration from the oropharynx rather than the stomach. This microbial data supports the clinical observation that antireflux surgery fails to prevents pulmonary complications such as pneumonias or hospitalizations [3, 4, 5, 6, 7]. By simultaneously sampling multiple sites per patient, we show that the lung and stomach microbiomes are highly variable across patients and determined primarily by patient rather than body site. Understanding the relationships between aerodigestive communities in aspirating and non-aspirating patients provides insight into the potential pathophysiology behind aspiration-related respiratory outcomes and suggests potential diagnostics and therapeutics for future investigation.

2.5 Declarations

2.5.1 Acknowledgments

2.5.2 Funding

2.5.3 Author contributions

R.R. designed the study, recruited patients, and performed the endoscopies. R.R. led patient recruitment. A.L. and S.I. assisted with patient recruitment. K.L. performed and interpreted the videofluoroscopic swallow studies. K.M. performed the bronchoscopies in this study. K.F. and S.S. performed the DNA isolation for 16S sequencing. C.D. processed and analyzed the microbiome data. C.D., E.A., and R.R. interpreted the results. C.D. and R.R. wrote the manuscript.

2.5.4 List of abbreviations

EGD: esophagogastroduodenoscopy

BAL: bronchoalveolar lavage

MII: multichannel intraluminal impedance

VFSS: videofluoroscopic swallow study

OTU: operational taxonomic unit

JSD: Jensen-Shannon distance

PERMANOVA: permutational multivariate analysis of variance

AUC: area under the ROC (receiver operating characteristic) curve

2.6 Tables

	Normal	Aspirators	Not tested	Total
BAL	33	33	36	102
Oropharyngeal swab	43	36	97	176
Gastric fluid	48	41	58	147
Stool			20	20

Table 2.1: Number of patient samples for each body site.

	Normal	Aspirators	Not tested	Total
BAL and oropharyngeal swab	23	25	25	73
BAL and gastric fluid	28	29	32	89
Gastric fluid and oropharyngeal swab	35	32	45	112
Stool and oropharyngeal swab			20	20

Table 2.2: Number of patients with multiple body sites sequenced.

Within people		Across people	p
Lung and oropharynx	more different	oropharynx	$< 1 \times 10^{-8}$
	not significant	lungs	0.8
Lung and gastric fluid	more similar	lungs	$< 1 \times 10^{-11}$
	more similar	gastric	$< 1 \times 10^{-8}$
Gastric and oropharyngeal	more similar	gastric	$< 1 \times 10^{-11}$
	not significant	oropharyngeal	0.07

Table 2.3: **Lung and gastric microbial communities are driven primarily by person rather than body site.** For each patient and each aerodigestive site, we compared the average JSD between that patient's site and all other patients' communities of that same site with the JSD between that patient's site and their other two aerodigestive sites. For example, the top row shows the comparisons between (1) the average JSD between a patient's oropharyngeal community and all other oropharyngeal communities and (2) the JSD between that patient's own oropharyngeal and lung communities. We subtracted each patient's between-sites JSD from their average between-patient JSD and calculated Wilcoxon signed-rank p-values using Python's `scipy.stats.wilcoxon` function.

Family	Genus	Non-aspirator	Aspirator	Difference
Flavobacteriaceae		8.7	48.0	39.3
Fusobacteriaceae	Fusobacterium	30.4	68.0	37.6
Micrococcaceae	Rothia	8.7	44.0	35.3
Veillonellaceae	Veillonella	26.1	60.0	33.9
Prevotellaceae		43.5	76.0	32.5
Porphyromonadaceae	Porphyromonas	39.1	68.0	28.9
Streptococcaceae	Streptococcus	13.0	40.0	27.0
Veillonellaceae	Centipeda	8.7	32.0	23.3
Prevotellaceae	Prevotella	17.4	36.0	18.6
Leptotrichiaceae	Streptobacillus	21.7	40.0	18.3
Fusobacteriaceae	Fusobacterium	17.4	32.0	14.6
Aerococcaceae	Abiotrophia	21.7	28.0	6.3

Table 2.4: **Prevalence of lung-oropharynx exchanged OTUs.** Prevalence is calculated as the percentage of patients who have the OTU present in both their lungs and oropharynx, calculated separately among aspirators (N = 25) and non-aspirators (N = 23). OTUs are ordered by their differential prevalence in aspirators relative to non-aspirators, and are labeled with their family- and genus-level taxonomies. Blank genus names indicate OTUs which were not annotated at the genus level. A similar table for the lung-gastric exchange OTUs can be found in Supplementary Table A-1.

Lung-oropharynx OTUs (12)	AUC	p	N (non-asp/asp)
Lung	0.63	0.29	33/33
Oropharyngeal	0.48	0.59	43/36
Concordance	0.66	0.19	23/25

Lung-gastric OTUs (74)			
Lung	0.63	0.19	33/33
Gastric fluid	0.66	0.04	48/41
Concordance	0.56	0.71	28/29

Table 2.5: **Classifiers based on the presence of exchanged OTUs.** (Top) Classifiers built from the presence of lung-oropharynx exchanged OTUs. (Bottom) Classifiers built from the presence of lung-gastric exchanged OTUs. Rows indicate which microbial community was used to train each classifier. In the "concordance" classifiers, OTUs which were either present or absent in both sites were coded as 1 and OTUs which were present in one site but absent in the other were coded as 0. AUCs are calculated as the area under the average ROC curve from five-fold cross validation. Fisher's exact p values are calculated on the predictions on the hold-out data for all cross validation folds. Each classifier was built 100 times with random patient splits and classifier initializations, and mean values are reported here. Similar classifiers built from the abundance of exchanged OTUs are shown in Supplementary Table A-2. AUCs and Fisher p-values from all 100 repetitions for all classifiers are shown in Supplementary Figures A-2 and A-3.)

Sites	AUC	Fisher p-value	N (non-asp/asp)
Lung	0.66	0.2	33/33
Oropharyngeal swab	0.71	0.02	43/36
Gastric fluid	0.67	0.11	48/41
Lung and oropharyngeal swab	0.81	0.01	23/25
Lung and gastric fluid	0.70	0.07	28/29
Oropharyngeal swab and gastric fluid	0.76	0.02	35/32
All three sites	0.83	0.01	19/23

Table 2.6: **Classifiers based on perturbed relationship between lung and oropharyngeal microbiota can distinguish aspirators from non-aspirators.** Areas under the ROC curve (AUC) and Fisher p-values calculated from classifiers trained on the entire microbial communities. Each row is a different classifier based on different combinations of aerodigestive communities, indicated in the "Sites" column. In the multi-site classifiers, the abundances of OTUs in different sites were used as separate features. For each classifier type, 100 classifiers were built, with random patient splits and classifier initializations. Mean values are reported. The distribution of AUCs and Fisher p-values from all 100 repetitions are shown in Supplementary Figure A-4.

	Mean (std)
Number of acid episodes	24.1 (26.9)
Number of nonacid episodes	14.8 (16.4)
Number of pH only episodes	16.6 (14.8)
Number of total reflux episodes	38.9 (33.8)
Percent time proximal reflux	0.005 (0.005)
Percent time distal reflux	0.012 (0.011)
Percent time pH < 4	7.2 (11.5)
Number abnormal by pH-metry	9
Number abnormal by MII	4

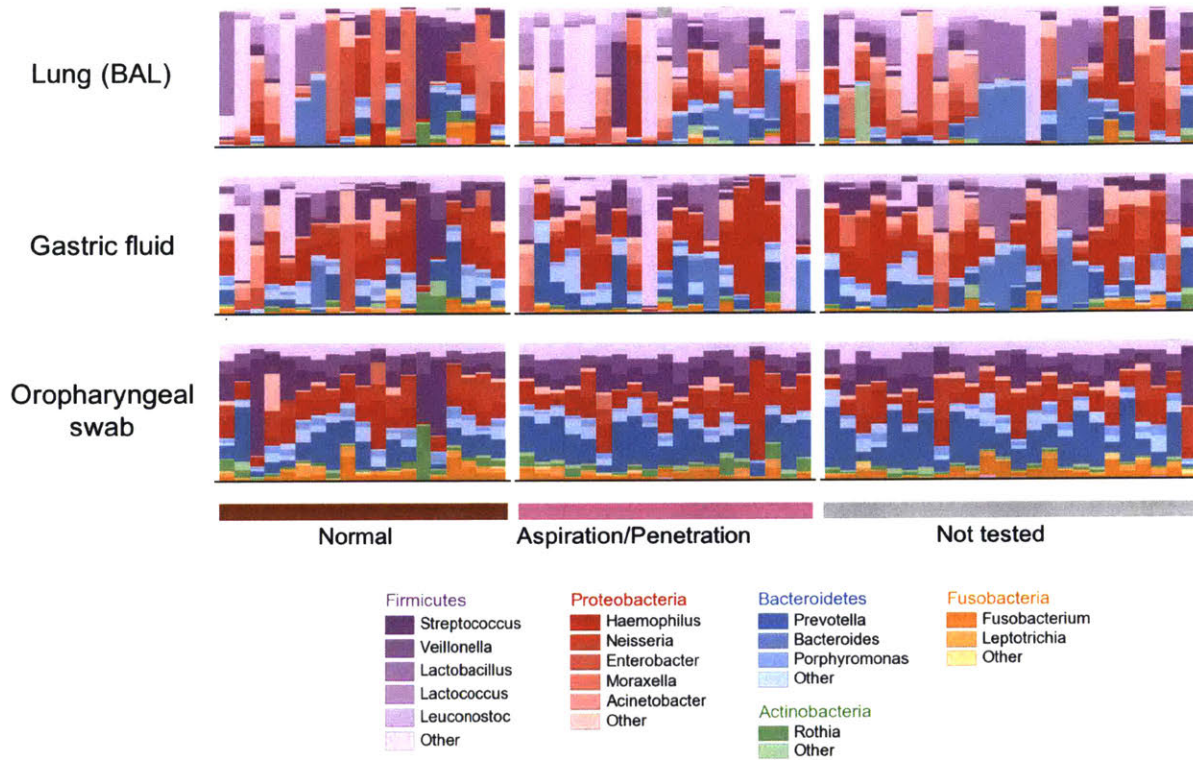Table 2.7: Reflux characteristics measured by pH-MII.

2.7 Figures

Figure 2-1: **Aerodigestive communities have similar predominant genera.** Bar plots showing relative abundances of aerodigestive OTUs collapsed to the genus level. Each column corresponds to one patient who had all three aerodigestive sites sequenced (N = 19 non-aspirators, 23 aspirators, 24 untested). Phyla in legend are those with mean abundance > 0.01 across all patients. Any other phyla are colored gray.
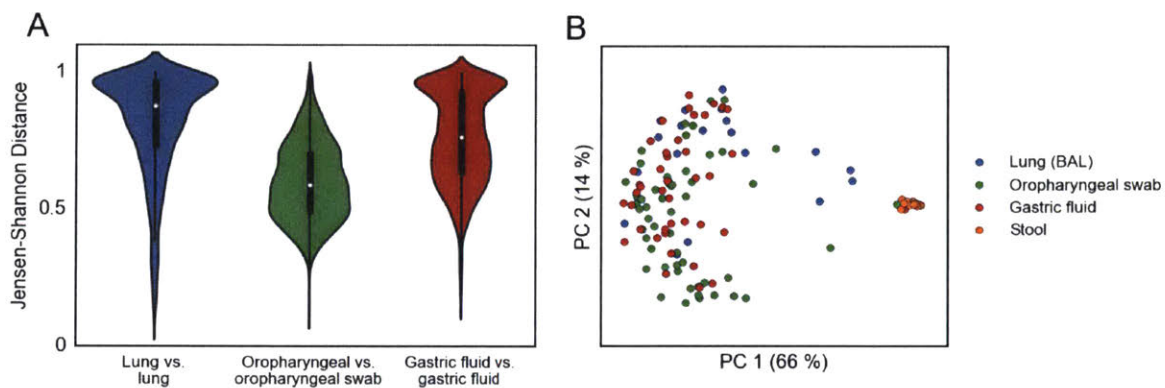
Figure 2-2: **Lung and gastric communities are more variable across people than oropharyngeal communities.** (A) Violin plot of the Jensen-Shannon distance (JSD) between samples from the same site across different patients. A JSD close to 1 indicates that communities are very different (less similar). (B) PCoA plots of aerodigestive and stool microbial communities for all patients in the 2016 sequencing batch (N = 21 BAL, 52 oropharyngeal swab, 43 gastric fluid, and 14 stool samples).
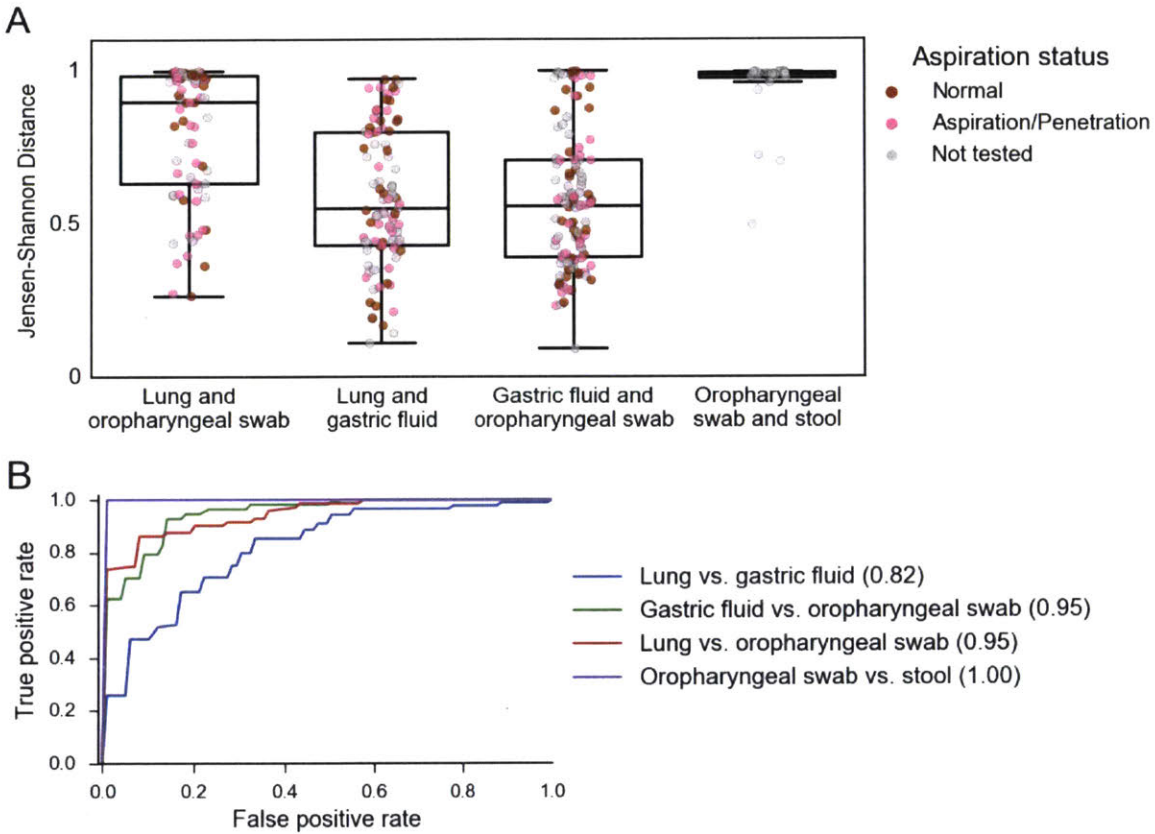
Figure 2-3: **Within patients, aerodigestive communities are similar but lung and oropharynx remain most distinct.** (A) Jensen-Shannon distances between samples from different sites from the same patient. Comparisons between stool and oropharynx are included to contextualize these results, as these are expected to be very different. All comparisons are significant (Wilcoxon rank sums test calculated with Python's `scipy.stats.ranksums` function) except the lung and gastric fluid vs. gastric fluid and oropharyngeal swab beta diversities ($p = 0.6$). Lung and oropharyngeal vs. oropharyngeal and stool, $p = 0.005$. All other comparisons: $p < 1 \times 10^{-8}$. (B) ROC curve of classifiers distinguishing different aerodigestive sites. Mean areas under the ROC curve (AUCs) are reported in parentheses in the legend.
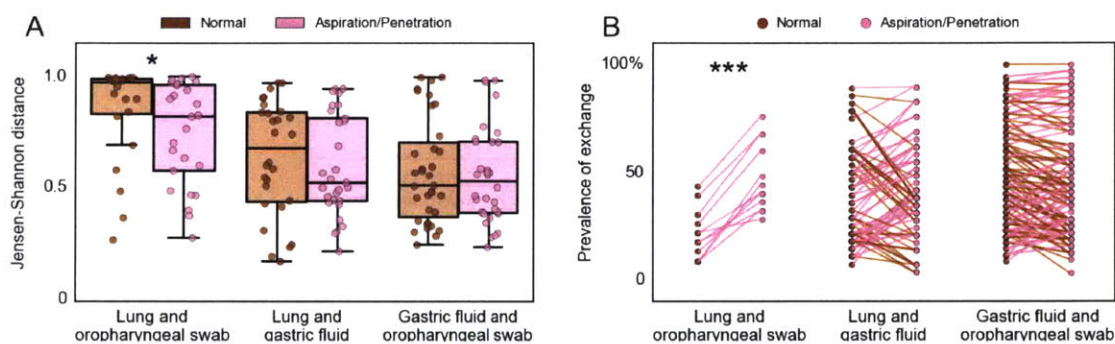
Figure 2-4: **Dysphagia increases aspiration of microbes from the oropharynx but not the stomach** (A) Intra-patient Jensen Shannon distance for different aerodigestive site comparisons in non-aspirators (brown) and aspirators (pink). Each point represents one patient. P-values (Wilcoxon rank sums test, calculated with Python's `scipy.stats.ranksums` function): lung and oropharyngeal swab $p = 0.04$, lung and gastric fluid $p = 0.5$, gastric fluid and oropharyngeal swab $p = 0.8$. (B) Percentage of patients with the previously defined exchanged microbes present in both of the respective sites (x-axis) in non-aspirators (brown) and aspirators (pink). Each pair of points represents one exchanged OTU. P-values (paired t-test on log_{10} prevalence values, calculated with Python's `scipy.stats.ttest_rel` function: lung and oropharyngeal swab $p = 3 \times 10^{-5}$, lung and gastric fluid $p = 0.8$, gastric fluid and oropharyngeal swab $p = 0.09$.
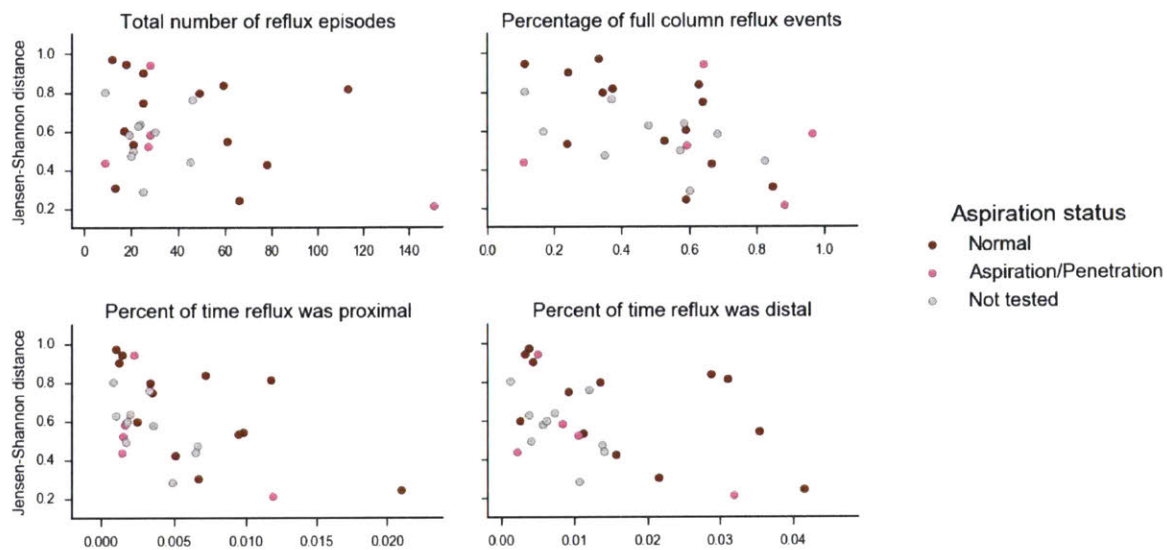
Figure 2-5: **Reflux severity may correlate with the similarity between lung and gastric communities.** Each plot shows the correlation between different reflux measures and the within-patient Jensen-Shannon distance between BAL and gastric fluid samples. Points are colored according to aspiration status. All reflux measures include both acid- and non-acid reflux. Spearman correlation and p-values: total number of reflux episodes $\rho_s = -0.14, p = 0.5$, percentage of full column reflux events $\rho_s = -0.41, p = 0.03$, percent of time reflux was proximal $\rho_s = -0.47, p = 0.01$, percent of time reflux was distal $\rho_s = -0.43, p = 0.02$.

Bibliography

[1] M. A. Holas, K. L. DePippo, and M. J. Reding. Aspiration and relative risk of medical complications following stroke. *Archives of Neurology*, 51(10):1051–1053, oct 1994. doi: 10.1001/archneur.1994.00540220099020. URL https://doi.org/10.1001/archneur.1994.00540220099020.

[2] J. Thomson, M. Hall, L. Ambroggio, B. Stone, R. Srivastava, S. S. Shah, and J. G. Berry. Aspiration and non-aspiration pneumonia in hospitalized children with neurologic impairment. *Pediatrics*, 137(2):e20151612–e20151612, jan 2016. doi: 10.1542/peds.2015-1612. URL https://doi.org/10.1542/peds.2015-1612.

[3] Douglas C. Barnhart, Matthew Hall, Sanjay Mahant, Adam B. Goldin, Jay G. Berry, Roger G. Faix, J. Michael Dean, and Rajendu Srivastava. Effectiveness of fundoplication at the time of gastrostomy in infants with neurological impairment. *JAMA Pediatrics*, 167(10):911, oct 2013. doi: 10.1001/jamapediatrics.2013.334. URL https://doi.org/10.1001/jamapediatrics.2013.334.

[4] Steven L. Lee, Hooman Shabatian, Jin-Wen Hsu, Harry Applebaum, and Philip I. Haigh. Hospital admissions for respiratory symptoms and failure to thrive before and after nissen fundoplication. *Journal of Pediatric Surgery*, 43(1):59–65, jan 2008. doi: 10.1016/j.jpedsurg.2007.09.020. URL https://doi.org/10.1016/j.jpedsurg.2007.09.020.

[5] A. B. Goldin, R. Sawin, K. D. Seidel, and D. R. Flum. Do antireflux operations decrease the rate of reflux-related hospitalizations in children? *Pediatrics*, 118(6):2326–2333, dec 2006. doi: 10.1542/peds.2006-2212. URL https://doi.org/10.1542/peds.2006-2212.

[6] Jennifer Yeh, Sharon A. McGrath-Morrow, and Joseph M. Collaco. Oxygen weaning after hospital discharge in children with bronchopulmonary dysplasia. *Pediatric Pulmonology*, 51(11):1206–1211, apr 2016. doi: 10.1002/ppul.23442. URL https://doi.org/10.1002/ppul.23442.

[7] R. Srivastava, E. C. Downey, M. O'Gorman, P. Feola, M. Samore, R. Holubkov, M. Mundorff, B. C. James, P. Rosenbaum, P. C. Young, and J. M. Dean. Impact of fundoplication versus gastrojejunal feeding tubes on mortality and in preventing aspiration pneumonia in young children with neurologic impairment who have gastroesophageal reflux disease. *PEDIATRICS*, 123(1):338–345, 2009. doi: 10.1542/peds.2007-1740. URL https://doi.org/10.1542/peds.2007-1740.

[8] Christine M. Bassis, John R. Erb-Downward, Robert P. Dickson, Christine M. Freeman, Thomas M. Schmidt, Vincent B. Young, James M. Beck, Jeffrey L. Curtis, and Gary B. Huffnagle. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. *mBio*, 6(2):e00037–15, 2015. doi: 10.1128/mbio.00037-15. URL https://doi.org/10.1128%2Fmbio.00037-15.

[9] Emily S. Charlson, Kyle Bittinger, Andrew R. Haas, Ayannah S. Fitzgerald, Ian Frank, Anjana Yadav, Frederic D. Bushman, and Ronald G. Collman. Topographical continuity of bacterial populations in the healthy human respiratory tract. *American Journal of Respiratory and Critical Care Medicine*, 184(8):957–963, 2011. doi: 10.1164/rccm.201104-0655oc. URL https://doi.org/10.1164%2Frccm.201104-0655oc.

[10] Rachel Rosen, Lan Hu, Janine Amirault, Umakanth Khatwa, Doyle V. Ward, and Andrew Onderdonk. 16s community profiling identifies proton pump inhibitor related differences in gastric, lung, and oropharyngeal microflora. *The Journal of Pediatrics*, 166(4):917–923, apr 2015. doi: 10.1016/j.jpeds.2014.12.067. URL https://doi.org/10.1016/j.jpeds.2014.12.067.

[11] Arvind Venkataraman, Christine M. Bassis, James M. Beck, Vincent B. Young, Jeffrey L. Curtis, Gary B. Huffnagle, and Thomas M. Schmidt. Application of a neutral community model to assess structuring of the human lung microbiome. *mBio*, 6(1):e02284–14, jan 2015. doi: 10.1128/mbio.02284-14. URL https://doi.org/10.1128%2Fmbio.02284-14.

[12] Leopoldo N Segal, Alexander V Alekseyenko, Jose C Clemente, Rohan Kulkarni, Benjamin Wu, Hao Chen, Kenneth I Berger, Roberta M Goldring, William N Rom, Martin J Blaser, and Michael D Weiden. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome*, 1(1):19, 2013. doi: 10.1186/2049-2618-1-19. URL https://doi.org/10.1186%2F2049-2618-1-19.

[13] John R. Erb-Downward, Deborah L. Thompson, Meilan K. Han, Christine M. Freeman, Lisa McCloskey, Lindsay A. Schmidt, Vincent B. Young, Galen B. Toews, Jeffrey L. Curtis, Baskaran Sundaram, Fernando J. Martinez, and Gary B. Huffnagle. Analysis of the lung microbiome in the "healthy" smoker and in COPD. *PLoS ONE*, 6(2):e16384, feb 2011. doi: 10.1371/journal.pone.0016384. URL https://doi.org/10.1371%2Fjournal.pone.0016384.

[14] Alison Morris, James M. Beck, Patrick D. Schloss, Thomas B. Campbell, Kristina Crothers, Jeffrey L. Curtis, Sonia C. Flores, Andrew P. Fontenot, Elodie Ghedin, Laurence Huang, Kathleen Jablonski, Eric Kleerup, Susan V. Lynch, Erica Sodergren, Homer Twigg, Vincent B. Young, Christine M. Bassis, Arvind Venkataraman, Thomas M. Schmidt, and George M. Weinstock. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *American Journal of Respiratory and Critical Care Medicine*, 187(10):1067–1075, may 2013.

doi: 10.1164/rccm.201210-1913oc. URL `https://doi.org/10.1164%2Frccm.201210-1913oc`.

[15] Hutsaya Tantipong, Chantana Morkchareonpong, Songyod Jaiyindee, and Visanu Thamlikitkul. Randomized controlled trial and meta-analysis of oral decontamination with 2% chlorhexidine solution for the prevention of ventilator-associated pneumonia. *Infection Control & Hospital Epidemiology*, 29(02):131–136, feb 2008. doi: 10.1086/526438. URL `https://doi.org/10.1086/526438`.

[16] Mirelle Koeman, Andre J. A. M. van der Ven, Eelko Hak, Hans C. A. Joore, Karin Kaasjager, Annemarie G. A. de Smet, Graham Ramsay, Tom P. J. Dormans, Leon P. H. J. Aarts, Ernst E. de Bel, Willem N. M. Hustinx, Ingeborg van der Tweel, Andy M. Hoepelman, and Marc J. M. Bonten. Oral decontamination with chlorhexidine reduces the incidence of ventilator-associated pneumonia. *American Journal of Respiratory and Critical Care Medicine*, 173 (12):1348–1355, jun 2006. doi: 10.1164/rccm.200505-820oc. URL `https://doi.org/10.1164/rccm.200505-820oc`.

[17] Hong bo Hu, Han ju Huang, Qiao ying Peng, Jia Lu, and Xing yun Lei. Prospective study of colonization and infection because of pseudomonas aeruginosa in mechanically ventilated patients at a neonatal intensive care unit in china. *American Journal of Infection Control*, 38(9):746–750, nov 2010. doi: 10.1016/j.ajic.2010.02.012. URL `https://doi.org/10.1016/j.ajic.2010.02.012`.

[18] Leopoldo N. Segal, Jose C. Clemente, Jun-Chieh J. Tsay, Sergei B. Koralov, Brian C. Keller, Benjamin G. Wu, Yonghua Li, Nan Shen, Elodie Ghedin, Alison Morris, Phillip Diaz, Laurence Huang, William R. Wikoff, Carles Ubeda, Alejandro Artacho, William N. Rom, Daniel H. Sterman, Ronald G. Collman, Martin J. Blaser, and Michael D. Weiden. Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a th17 phenotype. *Nature Microbiology*, 1(5), apr 2016. doi: 10.1038/nmicrobiol.2016.31. URL `https://doi.org/10.1038/nmicrobiol.2016.31`.

[19] Rachel Rosen and Samuel Nurko. The importance of multichannel intraluminal impedance in the evaluation of children with persistent respiratory symptoms. *The American Journal of Gastroenterology*, 99(12):2452–2458, dec 2004. doi: 10.1111/j.1572-0241.2004.40268.x. URL `https://doi.org/10.1111/j.1572-0241.2004.40268.x`.

[20] Qiong Wang, George M Garrity, James M Tiedje, and James R Cole. Naive bayesian classifier for rapid assignment of rrna sequences into the new bacterial taxonomy. *Applied and environmental microbiology*, 73(16):5261–5267, 2007.

[21] E. K. Costello, C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon, and R. Knight. Bacterial community variation in human body habitats across space and time. *Science*, 326(5960):1694–1697, nov 2009. doi: 10.1126/science.1177486. URL `https://doi.org/10.1126/science.1177486`.

[22] The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402):207–214, jun 2012. doi: 10.1038/nature11234. URL `https://doi.org/10.1038/nature11234`.

[23] C. A. Lozupone, J. Stombaugh, A. Gonzalez, G. Ackermann, D. Wendel, Y. Vazquez-Baeza, J. K. Jansson, J. I. Gordon, and R. Knight. Meta-analyses of studies of the human microbiota. *Genome Research*, 23(10):1704–1714, jul 2013. doi: 10.1101/gr.151803.112. URL `https://doi.org/10.1101/gr.151803.112`.

[24] Kelly Weir, Sandra McMahon, Linda Barry, Robert Ware, I Brent Masters, and Anne B. Chang. Oropharyngeal aspiration and pneumonia in children. *Pediatric Pulmonology*, 42(11):1024–1031, 2007. doi: 10.1002/ppul.20687. URL `https://doi.org/10.1002/ppul.20687`.

[25] O. Ortega, O. Sakwinska, S. Combremont, B. Berger, J. Sauser, C. Parra, S. Zarcero, J. Nart, S. Carrión, and P. Clavé. High prevalence of colonization of oral cavity by respiratory pathogens in frail older patients with oropharyngeal dysphagia. *Neurogastroenterology & Motility*, 27(12):1804–1816, sep 2015. doi: 10.1111/nmo.12690. URL `https://doi.org/10.1111/nmo.12690`.

[26] Kentaro Akata, Kazuhiro Yatera, Kei Yamasaki, Toshinori Kawanami, Keisuke Naito, Shingo Noguchi, Kazumasa Fukuda, Hiroshi Ishimoto, Hatsumi Taniguchi, and Hiroshi Mukae. The significance of oral streptococci in patients with pneumonia with risk factors for aspiration: the bacterial floral analysis of 16s ribosomal RNA gene using bronchoalveolar lavage fluid. *BMC Pulmonary Medicine*, 16(1), may 2016. doi: 10.1186/s12890-016-0235-z. URL `https://doi.org/10.1186/s12890-016-0235-z`.

[27] Shinya Kageyama, Toru Takeshita, Michiko Furuta, Mikiko Tomioka, Mikari Asakawa, Shino Suma, Kenji Takeuchi, Yukie Shibata, Yasuyuki Iwasa, and Yoshihisa Yamashita. Relationships of variations in the tongue microbiota and pneumonia mortality in nursing home residents. *The Journals of Gerontology: Series A*, oct 2017. doi: 10.1093/gerona/glx205. URL `https://doi.org/10.1093/gerona/glx205`.

Chapter 3

Meta-analysis of gut microbiome studies identifies disease-specific and shared responses

Claire Duvallet, Sean M Gibbons, Thomas Gurry, Rafael A Irizarry, and Eric J Alm.

The figures are at the end of the chapter. The Supplementary Information is in Appendix B.

Abstract

Hundreds of clinical studies have demonstrated associations between the human microbiome and disease, yet fundamental questions remain on how we can generalize this knowledge. Results from individual studies can be inconsistent and comparing published data is further complicated by a lack of standard processing and analysis methods. Here, we introduce the MicrobiomeHD database, which includes 28 published case-control gut microbiome studies spanning ten diseases. We perform a cross-disease meta-analysis of these studies using standardized methods. We find consistent patterns characterizing disease-associated microbiome changes. Some diseases are associated with over 50 genera, while most show only 10 to 15 genus-level changes. Some diseases are marked by the presence of potentially pathogenic microbes whereas others are characterized by a depletion of health-associated bacteria. Furthermore, we show that about half of genera associated with individual studies are bacteria which respond to more than one disease. Thus, many associations found in case-control studies are likely not disease-specific but rather part of a non-specific, shared response to health and disease.

3.1 Introduction

The human gastrointestinal tract digests food, absorbs nutrients, and plays important roles in maintaining metabolic homeostasis. The microbes residing in our gut harvest energy from the food we eat, train our immune system, break down xenobiotics and other foreign products, and release metabolites and hormones important for regulating our physiology [1, 2, 3]. Chemical signals from our microbiota can act locally within the gut, and can also have larger systemic effects (e.g. the 'gut-brain axis') [4, 5, 6].

Due to the physiological interplay between humans and our microbial communities, many diseases are hypothesized to be associated with shifts away from a "healthy" gut microbiome. These include metabolic disorders, inflammatory and auto-immune diseases, neurological conditions, and cancer, among others [1, 3, 7, 8, 9]. Certain gut-related conditions (e.g. obesity and inflammatory bowel disease) have been extensively studied in human cohorts and in animal experiments, where significant and sometimes causal microbial associations have been shown. These studies have spurred research into a number of complex diseases with unclear etiologies where a connection to the microbiome is suspected.

Overall, our current understanding of the precise relationships between the human gut microbiome and disease remains limited. Existing case-control studies often report finding disease-associated microbial "dysbiosis". However, the term "dysbiosis" is inconsistently and often vaguely defined, and can have a wide range of interpretations [10, 11]. Thus, we lack a comprehensive understanding of precisely how microbial communities and specific microbes within those communities cause, respond to, or contribute to disease. Are different diseases characterized by distinct shifts in the gut microbiome? Are some diseases marked by an invasion of pathogens whereas others show a depletion of beneficial bacteria? Can we identify microbial biomarkers for certain conditions, which are consistently enriched or depleted in a disease across many patient cohorts? Finally, are some bacteria part of a non-specific "healthy" or "diseased" microbiome and consistently associated with health or disease in general?

One approach to synthesize existing knowledge is to identify consistencies across

studies through a meta-analysis, which allows researchers to find and remove false positives and negatives that may obscure underlying biological patterns. However, prior meta-analyses of case-control gut microbiome studies have yielded mixed results and did not contextualize their findings across multiple diseases [12, 13, 14]. For some conditions like inflammatory bowel disease (IBD), an overall difference in the gut microbiota was found within several studies, but no individual microbes were consistently associated with IBD across studies [12]. For other conditions like obesity, multiple meta-analyses have found little to no difference in the gut microbiomes of obese and lean patients [12, 13, 14], even though the microbiome has been causally linked to obesity in mouse models [3, 15]. These meta-analyses have been limited by focusing on only one or two diseases, and thus do not extend their findings across a broader landscape of human disease to answer more general questions about overall patterns of disease-associated microbiome shifts.

In this paper, we collected 28 published case-control 16S amplicon sequencing gut microbiome datasets spanning ten different disease states. We acquired raw data and disease metadata for each study and systematically re-processed and re-analyzed the data. We investigated whether consistent and specific disease-associated changes in gut microbial communities could be identified across multiple studies of the same disease. Certain diseases (e.g. colorectal cancer (CRC)) are marked by an enrichment of disease-associated bacteria, while others (e.g. IBD) are characterized by a depletion of health-associated bacteria. Some conditions (e.g. diarrhea) exhibit large-scale community shifts with many associated microbes, while most show only a handful of associations. However, many associations are not specific to individual diseases but rather respond to multiple disease states. In most studies, the majority of the individual disease-associated microbes were part of this set of bacteria that respond non-specifically to healthy and diseased states. Thus, associations from individual case-control studies should be interpreted with caution, as these microbes may be indicative of a shared response to disease rather than part of disease-specific differences. Together, these findings reveal distinct categories of dysbioses which can inform the development of microbiome-based diagnostics and therapeutics.

51

3.2 Results

3.2.1 Most disease states show altered microbiomes

To answer questions about the reproducibility and generalizability of reported associations between the human microbiome and disease, we collected, re-processed, and re-analyzed raw data from a collection of microbiome datasets. We included studies with publicly available 16S amplicon sequencing data (i.e. FASTQ or FASTA) for stool samples from at least 15 case patients which also had associated disease metadata (i.e. case or control disease labels). Studies which exclusively focused on children under 5 years old were excluded from our analyses. We identified over 50 suitable case-control 16S datasets, of which 28 were successfully downloaded, processed, and included in a publicly available database which we called MicrobiomeHD [16]. Characteristics of these datasets, including sample sizes, diseases and conditions, and references, are shown in Table 3.1 and Supplementary Table B.1. For each downloaded study, we processed the raw sequencing data through our 16S processing pipeline (`https://github.com/thomasgurry/amplicon_sequencing_pipeline`) (see Supplementary Tables B.2 and B.3 for detailed data sources and processing methods). 100% denovo OTUs were assigned taxonomy with the RDP classifier [17] (c = 0.5), converted to relative abundances by dividing by total sample reads, and collapsed to the genus level. OTUs which were not assigned at the genus level were discarded. By collapsing data to the genus level, we lost the sensitivity to detect fine-scale differences in species or strain abundances across case and control groups, but we minimized certain batch effects that plague comparisons across studies. Thus, we took a course-grained approach to optimize our ability to compare data across studies at the expense of phylogenetic resolution.

We first asked whether reported associations between the gut microbiome and disease would be recapitulated once we controlled for processing and analysis approaches. To test whether the gut microbiome is altered in a variety of disease states, we built genus-level random forest classifiers to classify cases from controls within each study. We compared the resulting area under the Receiver Operating Characteristic (ROC)

curves (AUC) across studies (Figure 3-1A, Supplementary Figure B-1). We could classify cases from controls (AUC > 0.7) for at least one dataset for all diseases except arthritis and Parkinson's disease, which each only had one study. Notably, all diarrhea datasets (except Youngster et al. (2014) [18], which had only 4 distinct control patients and thus was not included in this analysis) had very high classifiability (AUC > 0.9). We successfully classified patients from controls (AUC > 0.7) in three out of four IBD studies and all four CRC studies, which is consistent with previous work showing that these patients can be readily distinguished from controls using supervised classification methods [12, 19, 20, 21]. Thus, the microbiome is indeed altered in many different diseases.

3.2.2 Loss of beneficial microbes or enrichment of pathogens

We next wondered whether the specific type of alteration was consistent across independent cohorts of patients with the same disease. We performed univariate tests on genus-level relative abundances for each dataset independently and compared results across studies (Kruskal-Wallis (KW) test with the Benjamini-Hochberg false discovery rate (FDR) correction [22]). Our re-analyses of the studies were largely consistent with the originally reported results. The same taxonomic groups showed similar trends as in the original publications, despite differences in data-processing methodologies (see Supplementary Note B.1.1 for a full comparison of our re-analysis with previously published results). Furthermore, we found that the disease-associated changes in the microbiome could be categorized into meaningful groups which provide insight into possible etiologies or therapeutic strategies for different types of disease.

In some diseases, microbiome shifts are dominated by an enrichment of a small number of "pathogenic" bacteria. In these cases, it is possible that the microbes play a causal role and that they could be targeted with narrow-spectrum antimicrobials. Colorectal cancer is characterized by such a shift, and we found significant agreement across three of the four CRC studies [8, 20, 21, 23] (Figures 3-1B and 3-2, genus labels in Supplementary Figure B-2). Dysbiosis associated with CRC is generally characterized by increased prevalence of the known pathogenic or pathogen-associated

Fusobacterium, *Porphyromonas*, *Peptostreptococcus*, *Parvimonas*, and *Enterobacter* genera (i.e. these genera were higher in CRC patients in 2 or more studies, Figures 3-2 and 3-3A, genus labels in Supplementary Figures B-2 and B-3). *Fusobacterium* is associated with a broad spectrum of human diseases and *Porphyromonas* is a known oral pathogen [24, 25].

By contrast, other disease-associated microbiome shifts are characterized by a depletion of health-associated bacteria in patients relative to controls. In these cases, probiotics that replace missing taxa may be a better treatment strategy than anti-microbials. Across our four IBD studies, patient microbiomes were dominated by a depletion of genera in patients relative to controls, especially butyrate-producing *Clostridiales* [19, 26, 27, 28] (Figures 3-1B and 3-2, genus labels in Supplementary Figure B-2). In particular, five genera from the *Ruminococcacaea* and *Lachnospiracaea* families were consistently depleted in IBD patients relative to controls in at least two studies (Figure 3-3A, genus labels in Supplementary Figure B-3). While not all genera within *Ruminococcacaea* and *Lachnospiracaea* are verified short-chain fatty acid (SCFA) producers, the dominant genera within these families are known to harbor genes for short chain fatty acid production [29] and are often associated with colonic health [30, 31, 32]. We found similar results when comparing Crohn's disease and ulcerative colitis patients to controls separately, without any consistent patterns across datasets that distinguished either IBD sub-type (Supplementary Note B.1.2; Supplementary Figures B-4 and B-5).

Some conditions are characterized by a broad restructuring of gut microbial communities. In these cases, full community restoration strategies like fecal microbiota transplants may be more appropriate. For example, diarrhea consistently results in large-scale rearrangements in the composition of the gut microbiome, which is likely reflective of reduced stool transit time (Figures 3-1 and 3-2). We saw many microbes consistently associated with both *Clostridium difficile* infection (CDI) and non-CDI diarrhea (Figures 3-2 and 3-3A) [18, 33, 34, 35]. In general, Proteobacteria increase in prevalence in patients with diarrhea, with a concomitant decrease in the relative abundances of Bacteroidetes and some Firmicutes. In particular, we see a reduction

in butyrate-producing Clostridia, including genera within *Ruminococcaceae* and *Lachnospiraceae* families, which have been associated with a healthy gut [36]. We also see an increase in prevalence of genera that contain organisms often associated with lower pH and higher oxygen levels of the upper-gut, like *Lactobacillaceae* and *Enterobacteriaceae*, in patients with diarrhea (Figure 3-3A) [37]. Additionally, both CDI and non-CDI diarrhea patients had lower alpha diversity, a measure of overall community structure, than healthy controls in all studies (Supplementary Figures B-6-B-8). Consistent with the CDI and non-CDI diarrheal studies, we also found that organisms associated with the upper gut, like *Lactobacillus* and *Enterobacteriaceae*, appear to be enriched in IBD patients, who can present with diarrheal symptoms (Supplementary Figure B-2) [37, 38]. IBD patients also tended to have lower alpha diversities than controls (Crohn's disease vs. controls in three studies, ulcerative colitis vs. controls in two studies; Supplementary Figures B-6–B-8), though this difference was less drastic than in the diarrheal studies where all patients had active diarrhea.

In some studies, confounding variables may drive associations. For example, there were no consistent differences between cases and controls across HIV studies because of demonstrated confounders [39, 40, 41] (Figure 3-2, 3-3A). As in the original Lozupone et al. (2013) [40] study, we found enrichment in *Prevotella, Catenibacterium, Dialister*, and *Desulfovibrio* in HIV-positive patients, in addition to 8 other genera (Figure 3-2 and Supplementary Figure B-2). We also found depletion of *Bacteroides, Odoribacter, Anaerostipes, Parasutterella*, and *Alistipes* in HIV-positive patients relative to controls. However, the Noguera-Julian et al. (2016) study showed that the genera that were significantly associated with HIV in the Lozupone paper were strongly associated with sexual behavior (e.g. men who have sex with men were associated with much higher *Prevotella* levels), and our re-analysis also found conflicting results between these two studies (Figure 3-2). Thus, there is no consensus on what genera are associated with HIV. Obesity is another example where confounding variables may drive microbiome alterations. Three recent meta-analyses found no reproducible obesity-associated microbiome shifts [12, 13, 14], which is consistent with our classification results where we were only able to accurately classify obese

and control patients in two out of five studies (Zhu et al. (2013) [1], Turnbaugh et al. (2009) [42]; Figure 3-1A). Our genus-level re-analysis did find a few consistent genus-level associations between lean and obese patients [1, 42, 43, 44, 45]. Two genera, *Roseburia* and *Mogibacterium*, were significantly enriched in obese individuals across two of the obesity studies (Figure 3-3A). Furthermore, *Anaerovorax*, *Oscillibacter*, *Pseudoflavonifractor*, and *Clostridium IV* were depleted in obese patients relative to controls in two of the studies. However, two of the five studies had no significant genus-level associations (q < 0.05), despite one having a large sample size (Zupancic et al. (2012) [44]). This suggests that confounding factors like diet may have given rise to certain associations found in our re-analysis and previously reported in the literature [14]. More studies that control for potential confounders, like host behavior and diet, will be required for diseases like obesity and HIV, where associations with the microbiome remain unclear. Finally, patients in case-control cohorts are frequently on other medications such as antibiotics which may confound disease-associated microbiome shifts. Six of our datasets included antibiotics metadata, and of these only one dataset (Schubert et al, 2014 [33]) had more than 5 controls who were on antibiotics. Thus, it is very likely that disease-associated genera in conditions which are often treated with antibiotics (e.g. diarrhea, IBD) are confounded with antibiotic usage. Future case control studies should focus on better separating treatment and disease variables by collecting detailed metadata on antibiotic and other medication usage, and perhaps also by recruiting controls undergoing a variety of treatments.

3.2.3 Shared vs. disease-specific microbial responses

Finally, we sought to determine whether a unified microbiome response to general health and disease could be identified. Previous studies have proposed that reduced alpha diversity is a reliable indicator of disease-associated dysbiosis [34, 42, 46]. In our re-analysis, we found no consistent reduction of alpha diversity in case patients, with the exception of diarrhea and perhaps IBD (Supplementary Figures B-6–B-8). These results are consistent with previous meta-analyses, which found inconsistent relationships between alpha diversity and disease and very small effect sizes in non-

diarrheal diseases [12, 13]. To further address the question of whether we could find a robust, generalized signal for diseased microbiomes regardless of the disease type, we built random forest classifiers to distinguish healthy patients from any type of case patient. The AUCs from these general healthy vs. disease classifiers correlated strongly with the original single-dataset classification results, indicating that there is indeed a general microbiome signal that can be identified even across different diseases (see Supplementary Note B.1.3 and Supplementary Figure B-9).

Having putatively shown the presence of a generalized microbial response to disease, we next sought to identify individual genera which respond non-specifically to health and disease. We considered a genus to be part of the non-specific, shared microbial response if it was significantly enriched or depleted ($q <= 0.05$) in at least one dataset from at least two different diseases (see Supplementary Note B.1.4 and Supplementary Figures B-10 and B-11 for further discussion on alternative definitions and statistical significance of shared response). We identified 24 health-associated genera and 20 disease-associated genera out of the 152 genera that were significant in at least one dataset (Figure 3-3A, genus labels in Supplementary Figure B-3). We also found 7 genera that were both health- and disease-associated (i.e. they were enriched in controls across at least two diseases, but were also depleted in controls in different comparisons across at least two diseases) (Figure 3-3A, black). Perhaps these genera represent bacteria disproportionately affected by confounders or technical artifacts. Alternatively, different species or strains within these genera may play alternate roles across diseases or community contexts, giving rise to variable responses at the genus level.

We identified distinct sub-groups of microbes within the *Bacteroidetes* and *Firmicutes* phyla which respond non-specifically to health and disease (Figure 3-3A). The order *Clostridiales* (specifically the *Lachnospiraceae* and *Ruminococcacaea* families) is associated with health across multiple diseases while the order *Lactobacillales* and family *Clostridiales Incertae Sedis XI* are associated with disease. The majority of the the non-specific responders in the order *Clostridiales* were associated with health, comprising the majority of all of the microbes which were non-specifically associated

57

with healthy patients (17 genera out of 24 total health-associated genera). All of five of the non-specific responders in the order *Lactobacillales* were enriched in case patients across multiple diseases. *Lactobacillales* genera are adapted to the lower pH of the upper gastrointestinal tract [37]. Perhaps the shared disease-associated taxa are indicators of shorter stool transit times and disruptions in the redox state and/or pH of the lower intestine, rather than specific pathogens. These non-specific responders are consistent with the results from a recent meta-analysis of six metagenomics datasets, which also found *Lactobacillales* and *Clostridiales* microbes among the most discriminative classification features across multiple studies [47]. Finally, we found that the order *Bacteroidales* is more mixed: two *Bacteroidales* genera were non-specifically associated with health, one with disease, and two with both health and disease.

A majority of bacterial associations within individual studies overlap with the shared response. For each dataset that had at least one significant (q < 0.05) association, we calculated the percent of associated genera which were also part of the non-specific response in the same direction (Figure 3-3B). Strikingly, the majority of microbial responses were not specific to individual diseases; on average, 51% of a dataset's genus-level associations were genera which were associated with more than one disease. In light of this finding, it is important that researchers performing future case-control studies consider whether an identified microbial association is truly specific to their disease of interest or is instead responding to a common symptom (e.g. diarrhea) or perhaps generally associated with health or sickness. Additionally, they can use the knowledge that many microbes respond non-specifically to disease to narrow lists of putative causal or diagnostic biomarkers to microbes which fall outside of the shared response and are thus more likely to be specific to the disease being studied. Researchers can access an updated list of shared microbial responders from this analysis at the MicrobiomeHD database [16], or they can curate their own lists by performing similar cross-disease meta-analyses.

Bacteria which are non-specifically associated with health are both ubiquitous and abundant across people, whereas bacteria which are non-specifically associated

with disease are abundant when present but are not ubiquitous. We calculated the average relative abundance (i.e. the total relative abundance across all patients divided by the number of patients with non-zero abundance) and ubiquity (i.e. the number of patients with non-zero abundance divided by the total number of patients) for each genus in the shared response. We found that health-associated genera were more ubiquitous than disease-associated ones, but not necessarily more abundant (Figure 3-3C). Thus, presence/absence of the non-specifically disease-associated genera appears to be a better indicator of disease-associated microbial shifts than changes in their relative abundances. However, a small subset of the non-specifically disease-associated genera were relatively ubiquitous across patients. Among the most ubiquitous were *Escherichia/Shigella* and *Streptococcus*. *Escherichia* includes common commensal strains, as well as pathogenic strains [48], and is frequently present in healthy people's guts as well as over-represented in sick patients. Genera within *Enterobacteriaceae*, *Lactobacillaceae*, and *Streptococcaceae* families are dominant in the upper gastrointestinal tract [37, 49] and are present in many people's stool at low frequency. These taxa likely become enriched with faster stool transit time (i.e. signatures of diarrhea) [37, 50].

3.2.4 Within and cross-disease meta-analysis improves interpretability

Identifying disease-specific and non-specific microbial responses required comparing studies both within and across multiple diseases. Multiple studies of the same disease were necessary to identify shifts consistently associated with individual diseases. We did not find consistent bacterial associations for conditions with fewer than four datasets (Figure 3-1, 3-3A). Within-disease meta-analysis also increased our ability to interpret the results from any one dataset. Despite few significant differences, some of these studies (e.g. Zhang et al. (2013) [51], Zhu et al. (2013) [1]) had high classifiability of patients vs. controls (AUC > 0.7, Figure 3-1A), indicating that there may be a disease-associated shift that was not detected by univariate comparisons.

However, because few other studies of the same disease were available for comparison, we could not confidently interpret the classification results beyond the reported AUC. For other studies with high AUCs but few univariate associations (e.g. Vincent et al. (2013) [34], Morgan et al. (2012) [27], Chen et al. (2012) [23]), our confidence that the high AUCs reflect true disease-associated differences increased because the high AUCs were consistent with other classifiers from the same disease type.

Meta-analysis identified potential false positives and false negatives across studies and conditions. For example, we found that reported associations between alpha diversity and disease within individual studies tended to lose significance when looking across studies, except in the case of diarrhea and perhaps IBD (Supplementary Figures B-6–B-8). Another example of a potential false positive was the association between *Prevotella* and disease. Autism [2], rheumatoid arthritis [52], and HIV [40, 41] have each been reported to be associated with *Prevotella*. For each of these diseases, the associations with *Prevotella* were weakly significant or complicated by confounding factors. In our more statistically conservative re-analysis, we found no association between autism or arthritis and *Prevotella*. As mentioned previously, in the case of HIV, the association with *Prevotella* was due to demographic factors unrelated to disease [39]. Regardless of whether shifts in *Prevotella* are truly biologically related to each studied disease state, it is clear that such shifts are not specific to one particular condition and should not be reported as putative disease-specific biomarkers. We also found that certain signals picked out by meta-analysis did not always hold within individual studies. For example, studies with small sample sizes often had few or no significant associations (e.g. Vincent et al. (2014) [34], Chen et al. (2012) [23], and Willing et al. (2009) [28]). Here, the fact that other studies analyzing the same diseases consistently found associations strengthens the hypothesis that the lack of microbiome-associated signal in these studies was due to low power rather than a lack of true signal. Because individual studies are plagued by low statistical power, confounding variables, and batch effects which can obscure biological signals, the identification of disease-specific and non-specific microbial associations will continue to improve as more datasets and diseases are included in future meta-analyses.

3.3 Discussion

Here, we report patterns of disease-associated shifts in the human gut microbiome which differ in their directionality (i.e. fraction of disease-enriched vs. disease-depleted genera) and extent (i.e. total number of genera that differ between cases and controls). Some diseases are characterized by an invasion of pathogenic or disease-associated bacteria (e.g. CRC), while others largely show a depletion of health-associated microbes (e.g. IBD). Diarrheal illnesses induce large-scale rearrangement of many members of the microbiota, whereas other conditions show fewer associations. We also find a set of microbes which are non-specifically associated with multiple diseases and show that these microbes comprise many of the disease-associated genera within any given study.

The identification of a non-specific microbial response is an important concept that should be considered in future case-control microbiome studies. It suggests that studies should be interpreted with extra caution, as many identified microbial associations may be indicative of a shared response to health or disease rather than a disease-specific biological difference. Microbes that are non-specifically associated with multiple diseases would not be useful as disease-specific diagnostics or to address causality [10]. On the other hand, bacteria that are associated with healthy patients across multiple diseases could be developed into a general probiotic which may be suited for many different conditions.

Additionally, characterizing "dysbioses" by their directionality and extent is a useful framework to generate hypotheses for future research on complex, heterogenous diseases with links to the microbiome. For example, the search for microbiome-based diagnostics may be more appropriate for diseases with consistently enriched disease-associated microbes, like CRC. On the other hand, patients with diseases which are characterized by depletion of health-associated microbes, like IBD, may benefit from prebiotic or probiotic interventions designed to enrich for these taxa. Furthermore, conditions which are characterized by large-scale shifts in community structure may be well-suited to treatment with fecal microbiota transplantation, as in CDI [18]. While

many of these conditions are unlikely to be fully treated by antibiotics, probiotics, or fecal microbiota transplants, our proposed framework could guide the search for new therapies and etiologies by generating testable hypotheses with higher likelihoods of success [10].

This analysis is the first to compare microbiome studies across more than two different diseases and highlights the importance of making raw data and associated patient metadata publicly available to enable future, more comprehensive analyses. This analysis does not include all possible studies, and certain important gastrointestinal diseases (e.g. irritable bowel syndrome) are missing, largely due to data and metadata availability. Future studies should expand on this work by including more cohorts from the same diseases as well as more diseases. To re-analyze these studies, we applied standard methods commonly used in the field and assumed that the original study designs and patient selection methods were adequate. We were reassured to find that a straightforward and standardized approach was able to recover very similar results to those previously reported in the various papers. Thus, we did not formally investigate heterogeneity between cohorts or technical inter-study batch effects. However, it is clear from our genus-level results that there is significant variation even across studies of the same disease. There are many possible reasons for this variation (experimental and sequencing artifacts, host-related covariates, stochastic disease-associated community changes etc. [11, 53, 54]), and future analyses should consider methods to correct for host confounders and technical batch effects. Concerns about batch effects motivated us to analyze the data at the genus level, which necessarily limited our resolution and biological interpretations of identified associations (e.g. different species or strains within a genus may have different associations with disease, which would not be captured in this analysis). Making raw data from case-control studies publicly available will also allow researchers to develop methods to correct for these batch effects, in addition to enabling more comprehensive future meta-analyses.

Despite the limitations of this study, our results provide more nuanced insight into dysbiosis, revealing distinct types of alterations that more precisely describe disease-

associated microbiome shifts. As the number of case-control cohorts increases, similar meta-analyses could be used to compare related diseases and identify microbiome alterations associated with general host physiological changes. For example, there may be a group of microbes which respond to or cause systemic inflammation. Could we identify these microbes by comparing multiple inflammatory or auto-immune diseases and study them to better understand the interactions between the microbiome and our immune system? Furthermore, some microbes may be consistently associated with neurological conditions and could contribute to the gastrointestinal symptoms that accompany or precede neurological manifestations [2, 9]. Studying these microbes could help us understand the 'gut-brain axis' by identifying common neuroactive molecules produced by these bacteria, which could also be used as targets for new treatments [4, 5, 6]. Finally, meta-analysis could be used to identify subsets of patients who exhibit distinct microbiome shifts within heterogenous diseases like IBD or in conditions which exhibit stochastic microbial responses, allowing for further stratification of disease subtypes and microbiome disruptions [11, 28, 55]. This work demonstrates that employing standard methods to contextualize new results within the broader landscape of clinically relevant microbiome studies is feasible and adds value to individual analyses. As excitement in this field grows, researchers should harness the increasing number of replicated case-control studies to swiftly and productively advance microbiome science from putative associations to transformative clinical impact.

3.4 Methods

3.4.1 Dataset collection

We identified case-control 16S studies from keyword searches in PubMed and by following references in meta-analyses and related case-control studies. We included studies with publicly available raw 16S data (fastq or fasta) and metadata indicating case or control status for each sample. Most data was downloaded from online

repositories (e.g. SRA) or links provided in the original publications, but some were acquired after personal communication with the authors (Supplementary Table B.3). We did not include any studies which required additional ethics committee approvals or authorizations for access (e.g. controlled dbGaP studies). In studies where multiple body sites were sampled or where multiple samples were taken per patient, we also required the respective metadata to include those metadata. We analyzed only stool 16S samples, and excluded studies with fewer than 15 case patients. In CRC studies with multiple control groups (e.g. healthy and non-CRC adenoma), only the healthy patients were used as controls for all of our comparisons. In studies with non-healthy controls (e.g. non-IBD patients), these patients were used as controls (as in the original papers). In the Schubert et al. CDI study [33], which had both CDI and non-CDI diarrheal patients, each group was used as an independent case group compared with controls. We also analyzed the NASH and obese patients from the Zhu et al. study [1] as independent case groups. When obesity studies reported body mass index instead of obesity status, we considered patients with BMI less than 25 as our control group and patients with BMI greater than 30 as the case group.

3.4.2 16S processing

Raw data were downloaded and processed through our in-house 16S processing pipeline (https://github.com/thomasgurry/amplicon_sequencing_pipeline). Data and metadata were acquired as described in Supplementary Table B.3. When needed, we de-multiplexed sequences by finding exact matches to the provided barcodes and trimmed primers with a maximum of 1 mismatch. In general, sequences were quality filtered by truncating at the first base with $Q < 25$. However, some datasets did not pass this stringent quality threshold (i.e. the resulting OTU table was either missing many of the original samples, or the read depth was significantly lower than reported in the original paper). For 454 data, we loosened the quality threshold to 20, whereas for paired-end Illumina data we removed reads with more than 2 expected errors. If possible, all reads were trimmed to 200 bp. In cases where this length trimming discarded a majority of sequences, we lowered our threshold to 150 or 101

bp. The specific processing parameters we used for each dataset can be found in Supplementary Table B.2. To assign OTUs, we clustered OTUs at 100% similarity using USEARCH [56] and assigned taxonomy to the resulting OTUs with the RDP classifier [17] and a confidence cutoff of 0.5. For each dataset, we removed samples with fewer than 100 reads and OTUs with fewer than 10 reads, as well as OTUs which were present in fewer than 1% of samples within a study. We calculated the relative abundance of each OTU by dividing its value by the total reads per sample. We then collapsed OTUs to genus level by summing their respective relative abundances, discarding any OTUs which were unannotated at the genus level. All statistical analyses were performed on this genus-level relative abundance data.

3.4.3 Statistical analyses

To perform supervised classification of cases and controls within each dataset, we built Random Forest classifiers with 5-fold cross-validation. To build our train and test sets, we used the python `scikit-learn StratifiedKFold` function with shuffling of the data [57]. To build our classifiers, we used the `RandomForestClassifier` function with 1000 estimators and other default settings [57]. We found no significant effect of various Random Forest parameters on the AUCs (Supplementary Figures B-12 and B-13). We calculated the interpolated area under the ROC curve (AUC) for each classifier based on the cross-validation testing results. To account for spurious high classifiability due to class imbalances, we also calculated the Cohen's kappa score for each classifier using `sklearn.metrics.cohen_kappa_score` on the test set predictions (Supplementary Table B.4). The kappa scores correlated well with the AUCs (Pearson $\rho = 0.9$), indicating that the majority of the classifiers performed well even when considering their underlying data distributions. We excluded Youngster et al. (2014) [18], which had only 4 distinct control patients, from all classifier analyses.

We performed univariate analyses on the relative abundances of genera in cases and controls with a non-parametric Kruskal-Wallis test using the `scipy.stats.mstats.kruskalwallis` function [58]. We corrected for multiple hypothesis testing in each dataset with the Benjamini-Hochberg false discovery rate

[22]. We performed all analyses on genus-level relative abundances for each dataset individually, and then compared these results across all studies.

We considered a genus to be consistently associated with a disease (Figure 3-3A, bottom) if it was significantly associated ($q < 0.05$) with the disease in the same direction in at least two studies of that disease. We considered a genus to be a non-specific microbial association (Figure 3-3A, top) if it was significantly associated ($q < 0.05$) in at least one dataset of at least two different diseases in the same direction. When we defined these non-specific genera, we did not include datasets which used non-healthy controls (Papa et al. (2012) [19] and Gevers et al. (2014) [26]) and the Lozupone et. al (2013) dataset [40], where the microbiome signal reflected behavior rather than disease state [39].

To build our generalized healthy vs. disease classifiers, we first concatenated metadata and genus-level abundance data for all datasets which had healthy controls (i.e. all datasets except Papa et al. (2012) [19] and Gevers et al. (2014) [26], which used non-IBD patients as controls, and CDI Youngster (2014), [18], which had only 4 distinct controls). We performed leave-one-dataset-out and leave-one-disease-out cross validation and calculated an AUC for each of the testing results.

3.4.4 Microbiome community analyses

Alpha diversities were calculated based on the non-collapsed 100% OTU-level relative abundances, and included un-annotated OTUs. We calculated alpha diversity metrics with the `skbio.math.diversity.alpha.chao1`, `shannon`, and `simpson` implementations.

We calculated the average abundance and ubiquity (Figure 3-3C) of each genus as the mean of its average values in each dataset across all patients. To calculate the abundance of each genus, we first calculated each genus's mean abundance within each dataset. We counted only patients with non-zero abundance of the genus in this calculation. We then took the average of these mean abundances across all datatsets. To calculate the ubiquity of each genus, we calculated the percent of patients with non-zero abundance of that genus in each dataset. We then took the average of these

66

ubiquities across all datasets.

3.4.5 Code availability

The code to reproduce all of the analyses in this paper is available at `https://github.com/cduvallet/microbiomeHD`. We encourage researchers to incorporate their existing and future case-control studies into the MicrobiomeHD database by contacting us.

3.4.6 Data availability

Raw sequencing data for each study can be accessed as described in Supplementary Table B.3. The raw processed OTU tables can be accessed at the MicrobiomeHD database, available at `https://doi.org/10.5281/zenodo.840333` [16].

Supplementary files, including the q-values for all genus-level comparisons in every dataset, disease-associated genera for the diseases with more than three datasets, and a list of non-specific genera are also available at `https://github.com/cduvallet/microbiomeHD`. All other relevant data supporting the findings of the study are available in this article and its Supplementary Information files, or from the corresponding author upon request.

3.5 Acknowledgements

3.6 Author contributions

C.D., T.G., and E.A. conceived of the research. C.D., S.G., and T.G. identified and downloaded the datasets. C.D. and T.G. wrote the code to process data. C.D. processed the data and performed all analyses. C.D., S.G, R.I., and E.A. interpreted the results and prepared the manuscript.

3.6.1 Competing Financial Interests

E.A. is on the Board of Directors of OpenBiome, a non-profit stool bank. E.A. is a co-founder of Finch Therapeutics, which aims to develop microbiome-based therapeutics. The remaining authors declare no competing financial interests.

3.7 Tables

Dataset ID	Controls	N (controls)	Cases	N (cases)	Ref.
Singh 2015, EDD	H	82	EDD	201	[35]
Schubert 2014, CDI	H	154	CDI	93	[33]
Schubert 2014, nonCDI	H	154	nonCDI	89	[33]
Vincent 2013, CDI	H	25	CDI	25	[34]
Youngster 2014, CDI	H	4	CDI	19	[18]
Goodrich 2014, OB	H	428	OB	185	[43]
Turnbaugh 2009, OB	H	61	OB	195	[42]
Zupancic 2012, OB	H	96	OB	101	[44]
Ross 2015, OB	H	26	OB	37	[45]
Zhu 2013, OB	H	16	OB	25	[1]
Baxter 2016, CRC	H	172	CRC	120	[20]
Zeller 2014, CRC	H	75	CRC	41	[21]
Wang 2012, CRC	H	54	CRC	44	[8]
Chen 2012, CRC	H	22	CRC	21	[23]
Gevers 2014, IBD	nonIBD	16	CD	146	[26]
Morgan 2012, IBD	H	18	UC, CD	108	[27]
Papa 2012, IBD	nonIBD	24	UC, CD	66	[19]
Willing 2010, IBD	H	35	UC, CD	45	[28]
Noguera-Julian 2016, HIV	H	34	HIV	205	[39]
Dinh 2015, HIV	H	15	HIV	21	[41]
Lozupone 2013, HIV	H	13	HIV	23	[40]
Son 2015, ASD	H	44	ASD	59	[7]
Kang 2013, ASD	H	20	ASD	19	[2]
Alkanani 2015, T1D	H	55	T1D	57	[59]
Mejia-Leon 2014, T1D	H	8	T1D	21	[60]
Wong 2013, NASH	H	22	NASH	16	[61]
Zhu 2013, NASH	H	16	NASH	22	[1]
Scher 2013, ART	H	28	PSA, RA	86	[52]
Zhang 2013, LIV	H	25	CIRR, MHE	46	[51]
Scheperjans 2015, PAR	H	74	PAR	74	[9]

Table 3.1: Datasets collected and processed through standardized pipeline. Disease labels: ART = arthritis, ASD = austism spectrum disorder, CD = Crohn's disease, CDI = *Clostridium difficile* infection, CIRR = liver cirrhosis, CRC = colorectal cancer, EDD = enteric diarrheal disease, H = healthy, HIV = human immunodeficiency virus, LIV = liver diseases, MHE = minimal hepatic encephalopathy, NASH = non-alcoholic steatohepatitis, OB = obesity, PAR = Parkinson's disease, PSA = psoriatic arthritis, RA = rheumatoid arthritis, T1D = type I diabetes, UC = ulcerative colitis. nonCDI controls are patients with diarrhea who tested negative for *C. difficile* infection. nonIBD controls are patients with gastrointestinal symptoms but no intestinal inflammation. Datasets are ordered as in Figure 3-1.
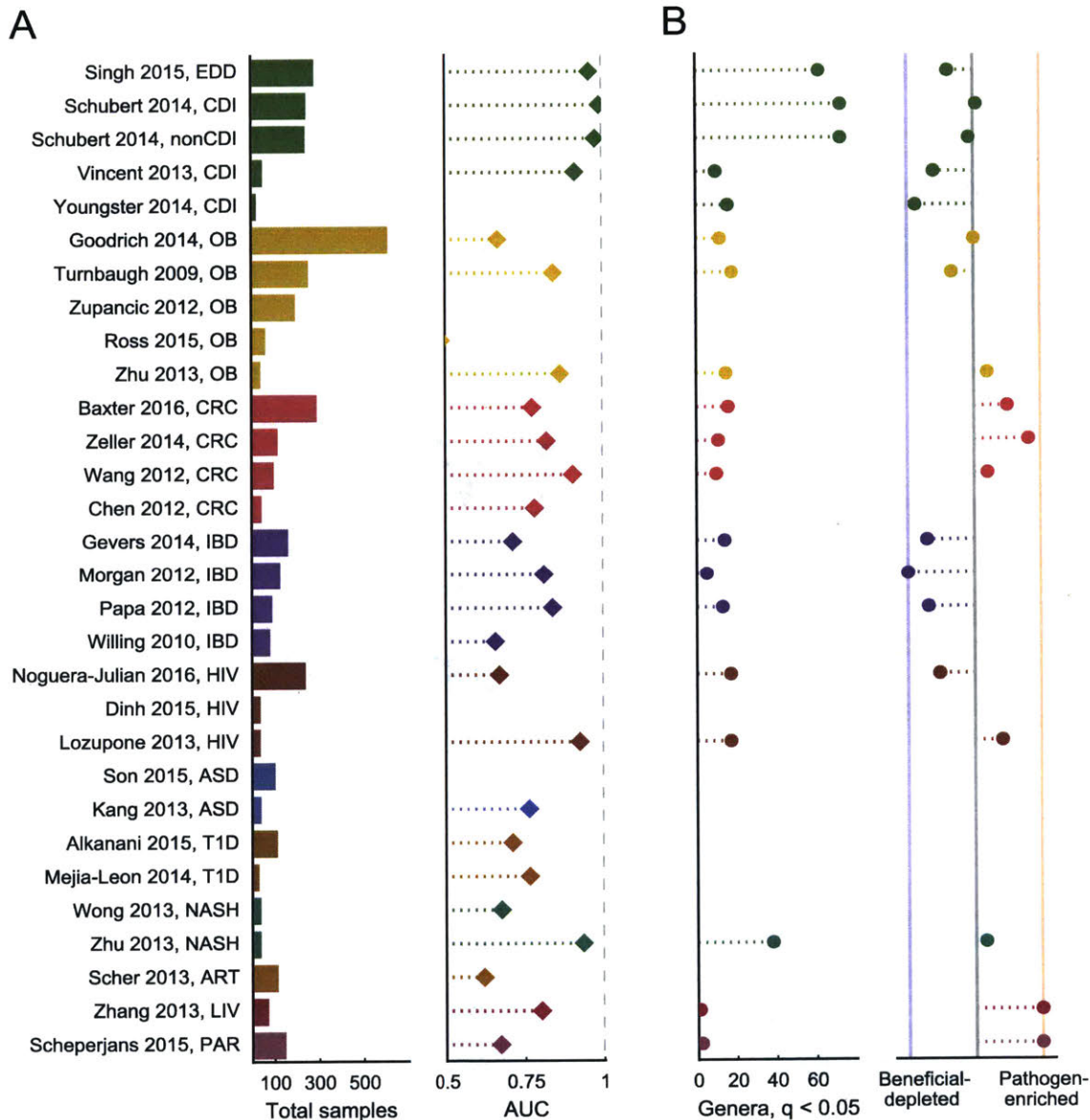
3.8 Figures

Figure 3-1: **Most diseases show microbiome alterations, and consistent disease-associated shifts differ in their extent and direction.** (A) Left: Total sample size for each study included in these analyses. Additional information about each dataset can be found in Table 3.1. Studies on the y-axis are grouped by disease and ordered by decreasing sample size (top to bottom). Right: Area under the ROC curve (AUC) for genus-level random forest classifiers. X-axis starts at 0.5, the expected value for a classifier which assigns labels randomly, and AUCs less than 0.5 are not shown. ROC curves for all datasets are in Supplementary Figure B-1. Note that Youngster et al. (2014) [18] had only 4 distinct control patients was excluded from the Random Forest analysis. (B) Left: Number of genera with q < 0.05 (Kruskal-Wallis (KW) test, Benjamini-Hochberg FDR correction) for each dataset. If a study has no significant associations, no point is shown. Right: Direction of the microbiome shift, i.e. the percent of total associated genera which were enriched in diseased patients. In datasets on the leftmost blue line, 100% of associated (q < 0.05, FDR KW test) genera are health-associated (i.e. depleted in patients relative to controls). In datasets on the rightmost red line, 100% of associated (q < 0.05, FDR KW test) genera are disease-associated (i.e. enriched in patients relative to controls). Supplementary Figures B-14 and B-15 show q-values and effects for each genus in each study.
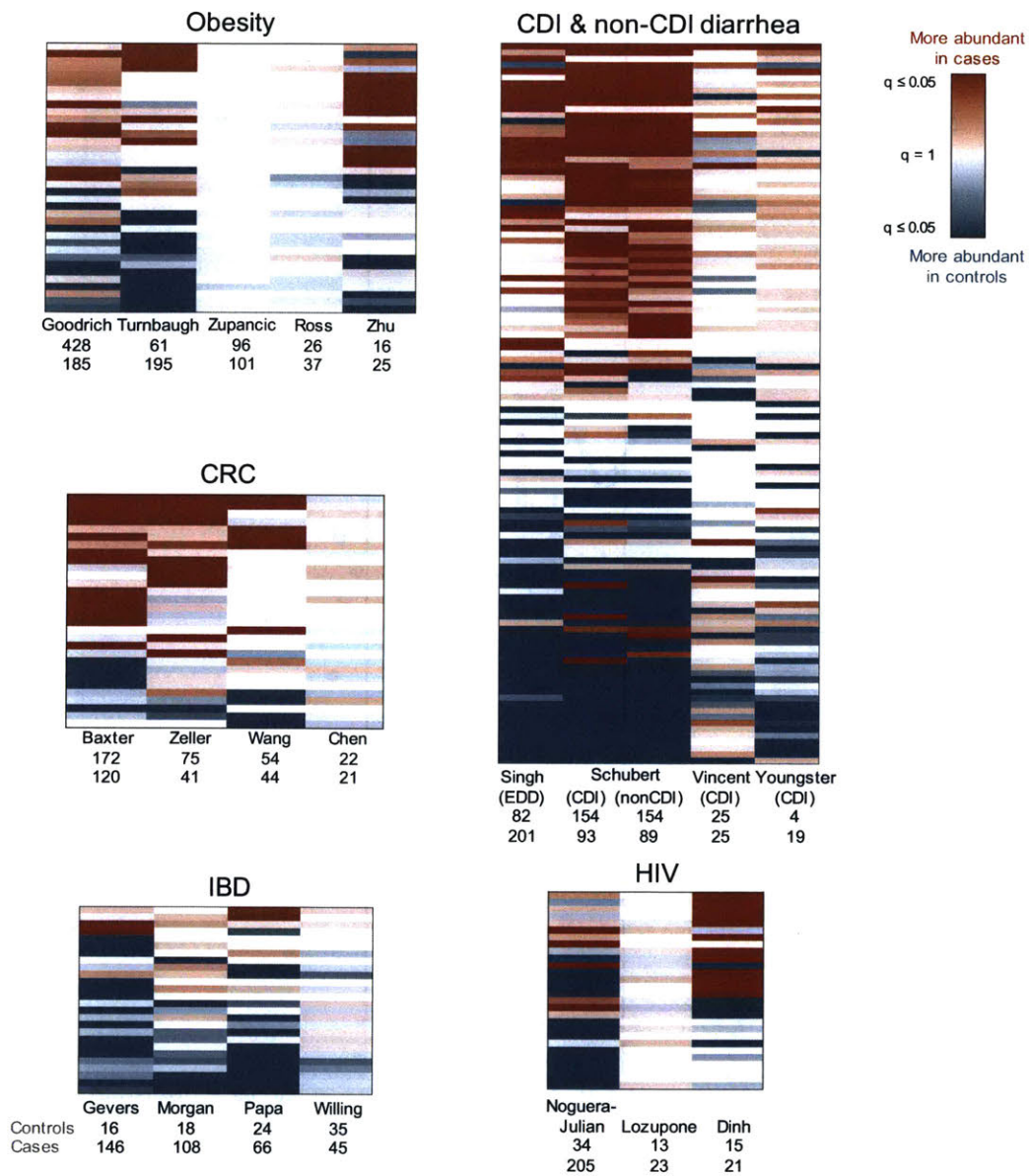
71

Obesity

Goodrich	Turnbaugh	Zupancic	Ross	Zhu
428	61	96	26	16
185	195	101	37	25

CDI & non-CDI diarrhea

Singh (EDD)	Schubert (CDI)	(nonCDI)	Vincent (CDI)	Youngster (CDI)
82	154	154	25	4
201	93	89	25	19

CRC

Baxter	Zeller	Wang	Chen
172	75	54	22
120	41	44	21

IBD

	Gevers	Morgan	Papa	Willing
Controls	16	18	24	35
Cases	146	108	66	45

HIV

Noguera-Julian	Lozupone	Dinh
34	13	15
205	23	21

More abundant in cases
q ≤ 0.05
q = 1
q ≤ 0.05
More abundant in controls

Figure 3-2: **Comparing results from multiple studies of the same disease reveals patterns in disease-associated microbiome alterations.** Heatmaps showing log10(q-values) for each disease (Kruskal-Wallis (KW) test, Benjamini-Hochberg FDR correction). Rows include all genera which were significant in at least one dataset within each disease, columns are datasets. Q-values are colored by direction of the effect, where red indicates higher mean abundance in disease patients and blue indicates higher mean abundance in controls. Opacity ranges from q = 0.05 to 1, where q values less than 0.05 are the most opaque and q values close to 1 are gray. White indicates that the genus was not present in that dataset. Within each heatmap, rows are ordered from most disease-associated (top) to most health-associated (bottom) (i.e. by the sum across rows of the log10(q-values), signed according to directionality of the effect). The extent of a disease-associated microbiome shift can be visualized by the number of rows in each disease heatmap; the directionality of a shift can be seen in the ratio of red rows to blue rows within each disease. See Supplementary Figure B-2 for genus (row) labels.
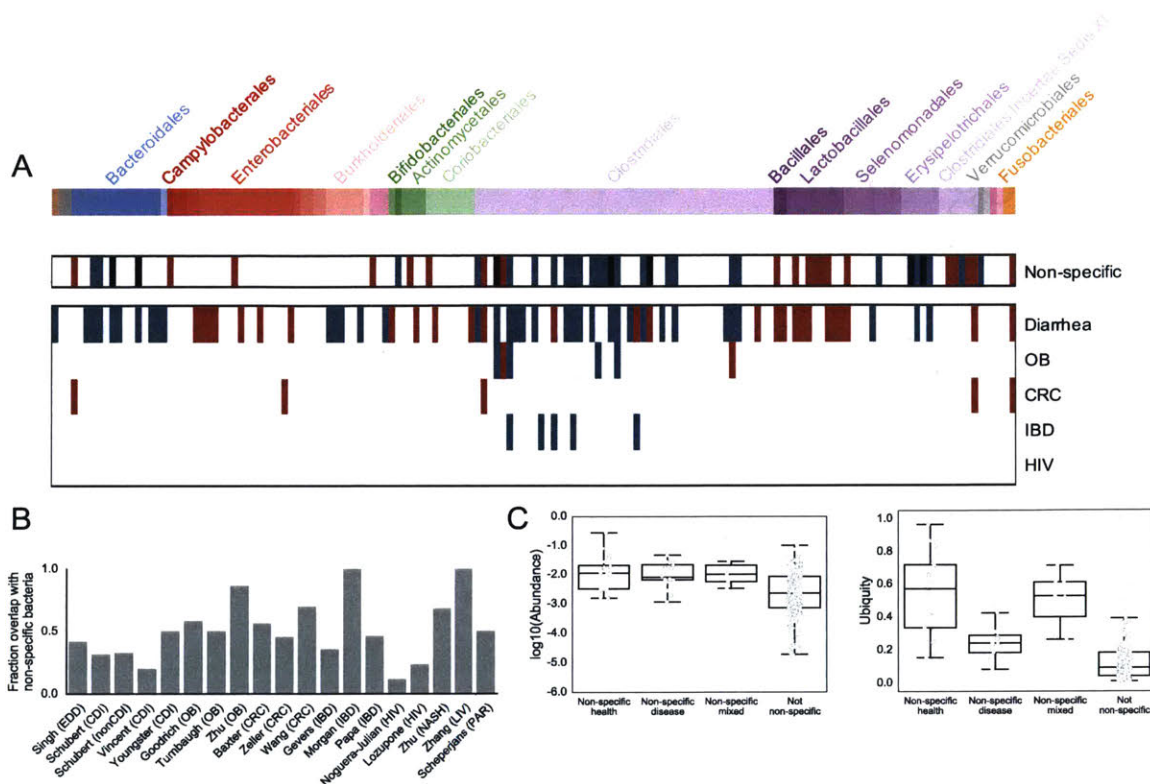
Figure 3-3: **The majority of disease-associated microbiome alterations overlap with a non-specific microbial response to disease.** (A) Non-specific and disease-associated genera. Genera are in columns, arranged phylogenetically according to a PhyloT tree built from genus-level NCBI IDs (`http://phylot.biobyte.de`). Non-specific genera are associated with health (or disease) in at least two different *diseases* (q < 0.05, Kruskal-Wallis (KW) test, Benjamini-Hochberg FDR correction). Disease-specific genera are significant in the same direction in at least two *studies* of the same disease (q < 0.05, FDR KW test). As in Figure 3-2, blue indicates higher mean abundance in controls and red indicates higher mean abundance in patients. Black bars indicate mixed genera which were associated with health in two diseases and also associated with disease in two diseases. Disease-specific genera are shown for diseases with at least 3 studies. Phyla, left to right: Euryarchaeota (brown), Verrucomicrobia Subdivision 5 (gray), Candidatus Saccharibacteria (gray), Bacteroidetes (blue), Proteobacteria (red), Synergistetes (pink), Actinobacteria (green), Firmicutes (purple), Verrucomicrobia (gray), Lentisphaerae (pink), Fusobacteria (orange). See Supplementary Figure B-3 for genus labels. (B) The percent of each study's genus-level associations which overlap with the shared response (q < 0.05, FDR KW test). Only datasets with at least one significant association are shown. (C) Overall abundance and ubiquity of non-specific genera across all patients in all datasets. Non-specific genera on the x-axis are as defined above.

Bibliography

[1] L. Zhu, S.S. Baker, C. Gill, W. Liu, R. Alkhouri, R.D. Baker, and S.R. Gill. Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH) patients: a connection between endogenous alcohol and NASH. *Hepatology*, 57(2):601–609, 2013.

[2] D.W. Kang, J.G. Park, Z.E. Ilhan, G. Wallstrom, J. LaBaer, J.B. Adams, and R. Krajmalnik-Brown. Reduced incidence of Prevotella and other fermenters in intestinal microflora of autistic children. *PloS one*, 8(7):e68322, 2013.

[3] Peter J Turnbaugh, Ruth E Ley, Michael A Mahowald, Vincent Magrini, Elaine R Mardis, and Jeffrey I Gordon. An obesity-associated gut microbiome with increased capacity for energy harvest. *nature*, 444(7122):1027–131, 2006.

[4] Elaine Y. Hsiao, Sara W. McBride, Sophia Hsien, Gil Sharon, Embriette R. Hyde, Tyler McCue, Julian A. Codelli, Janet Chow, Sarah E. Reisman, Joseph F. Petrosino, Paul H. Patterson, and Sarkis K. Mazmanian. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell*, 155(7):1451–1463, 2013.

[5] John Cryan and Timothy Dinan. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nature Reviews Neuroscience*, 13:701–712, 2012.

[6] Theofilos Poutahidis, Sean M. Kearney, Tatiana Levkovich, Peimin Qi, Bernard J. Varian, Jessica R. Lakritz, Yassin M. Ibrahim, Antonis Chatzigiagkos, Eric J. Alm, and Susan E. Erdman. Microbial symbionts accelerate wound healing via the neuropeptide hormone oxytocin. *PLoS ONE*, 8(10):e78898, 2013.

[7] J. Son, L.J. Zheng, L.M. Rowehl, X. Tian, Y. Zhang, W. Zhu, L. Litcher-Kelly, K.D. Gadow, G. Gathungu, C.E. Robertson, D. Ir, D.N. Frank, and E. Li. Comparison of fecal microbiota in children with autism spectrum disorders and neurotypical siblings in the simons simplex collection. *PLOS ONE*, 10(10):e0137725, 2015.

[8] T. Wang, G. Cai, Y. Qiu, N. Fei, M. Zhang, X. Pang, W. Jia, S. Cai, and L. Zhao. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *The ISME Journal*, 6(2):320–329, 2011.

74

[9] F. Scheperjans, V. Aho, P.A.B. Pereira, K. Koskinen, L. Paulin, E. Pekkonen, E. Haapaniemi, S. Kaakkola, J. Eerola-Rautio, P. Pohja, E. Kinnunen, K. Murros, and P. Auvinen. Gut microbiota are related to parkinson's disease and clinical phenotype. *Movement Disorders*, 30(3):350–358, dec 2014.

[10] Scott W Olesen and Eric J Alm. Dysbiosis is not an answer. *Nature Microbiology*, 1:16228, 2016.

[11] Jesse R. Zaneveld, Ryan McMinds, and Rebecca Vega Thurber. Stress and stability: applying the anna karenina principle to animal microbiomes. *Nature Microbiology*, 2(9):17121, Aug 2017.

[12] William A Walters, Zech Xu, and Rob Knight. Meta-analyses of human gut microbes associated with obesity and ibd. *FEBS letters*, 588(22):4223–4233, 2014.

[13] Marc A. Sze and Patrick D. Schloss. Looking for a signal in the noise: Revisiting obesity and the microbiome. *mBio*, 7(4), 2016.

[14] Mariel M. Finucane, Thomas J. Sharpton, Timothy J. Laurent, and Katherine S. Pollard. A taxonomic signature of obesity in the microbiome? getting to the guts of the matter. *PLoS ONE*, 2014.

[15] Vanessa K Ridaura, Jeremiah J Faith, Federico E Rey, Jiye Cheng, Alexis E Duncan, Andrew L Kau, Nicholas W Griffin, Vincent Lombard, Bernard Henrissat, James R Bain, et al. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science*, 341(6150):1241214, 2013.

[16] Claire Duvallet, Sean Gibbons, Thomas Gurry, Rafael Irizarry, and Eric Alm. MicrobiomeHD: The human gut microbiome in health and disease, 2017.

[17] Qiong Wang, George M Garrity, James M Tiedje, and James R Cole. Naive bayesian classifier for rapid assignment of rrna sequences into the new bacterial taxonomy. *Applied and environmental microbiology*, 73(16):5261–5267, 2007.

[18] I. Youngster, J. Sauk, C. Pindar, R.G. Wilson, J.L. Kaplan, M.B. Smith, E.J. Alm, D. Gevers, G.H. Russell, and E.L. Hohmann. Fecal microbiota transplant for relapsing clostridium difficile infection using a frozen inoculum from unrelated donors: A randomized, open-label, controlled pilot study. *Clinical Infectious Diseases*, 58(11):1515–1522, 2014.

[19] E. Papa, M. Docktor, C. Smillie, S. Weber, S.P. Preheim, D. Gevers, G. Giannoukos, D. Ciulla, D. Tabbaa, J. Ingram, D.B. Schauer, D.V. Ward, J.R. Korzenik, R.J. Xavier, A. Bousvaros, and E.J. Alm. Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease. *PLoS ONE*, 7(6):e39242, 2012.

[20] N.T. Baxter, M.T. Ruffin, M.A. Rogers, and P.D. Schloss. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Medicine*, 8(1), 2016.

[21] G. Zeller, J. Tap, A.Y. Voigt, S. Sunagawa, J.R. Kultima, P.I. Costea, A. Amiot, J. Bohm, F. Brunetti, N. Habermann, R. Hercog, M. Koch, A. Luciani, D.R. Mende, M.A. Schneider, P. Schrotz-King, C. Tournigand, J.T. Nhieu, T. Ya- mada, J. Zimmermann, V. Benes, M. Kloor, C.M. Ulrich, M. von Knebel Doe- beritz, I. Sobhani, and P. Bork. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Molecular Systems Biology*, 10(11):766–766, 2014.

[22] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a prac- tical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B Methodological)*, 57:289–300, 1995.

[23] W. Chen, F. Liu, Z. Ling, X. Tong, and C. Xiang. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS ONE*, 7(6):e39743, 2012.

[24] Yiping W Han. Fusobacterium nucleatum: a commensal-turned pathogen. *Cur- rent Opinion in Microbiology*, 23:141–147, 2015.

[25] Kaitlin J. Flynn, Nielson T. Baxter, and Patrick D. Schloss. Metabolic and community synergy of oral bacteria in colorectal cancer. *mSphere*, 1(3):e00102– 16, 2016.

[26] D. Gevers, S. Kugathasan, L.A. Denson, Y. Vázquez-Baeza, W. Van Treuren, B. Ren, E. Schwager, D. Knights, S. Song, M. Yassour, X.C. Morgan, A.D. Kos- tic, C. Luo, A. González, D. McDonald, Y. Haberman, T. Walters, S. Baker, J. Rosh, M. Stephens, M. Heyman, J. Markowitz, R. Baldassano, A. Grif- fiths, F. Sylvester, D. Mack, S. Kim, W. Crandall, J. Hyams, C. Huttenhower, R. Knight, and R. Xavier. The treatment-naive microbiome in new-onset crohn's disease. *Cell Host & Microbe*, 15(3):382–392, mar 2014.

[27] X.C. Morgan, T.L. Tickle, H. Sokol, D. Gevers, K.L. Devaney, D.V. Ward, J.A. Reyes, S.A. Shah, N. LeLeiko, S.B. Snapper, A. Bousvaros, J. Korzenik, B.E. Sands, R.J. Xavier, and C. Huttenhower. Dysfunction of the intestinal micro- biome in inflammatory bowel disease and treatment. *Genome Biol*, 13(9):R79, 2012.

[28] B.P. Willing, J. Dicksved, J. Halfvarson, A.F. Andersson, M. Lucio, Z. Zheng, G. J"arnerot, C. Tysk, J.K. Jansson, and L. Engstrand. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology*, 139(6):1844–1854.e1, dec 2010.

[29] Petra Louis, Pauline Young, Grietje Holtrop, and Harry J Flint. Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl- coa: acetate coa-transferase gene. *Environmental microbiology*, 12(2):304–314, 2010.

[30] Harry J Flint, Karen P Scott, Petra Louis, and Sylvia H Duncan. The role of the gut microbiota in nutrition and health. *Nature Reviews Gastroenterology and Hepatology*, 9(10):577–589, 2012.

[31] S Miquel, R Martin, O Rossi, LG Bermudez-Humaran, JM Chatel, H Sokol, M Thomas, JM Wells, and P Langella. Faecalibacterium prausnitzii and human intestinal health. *Current opinion in microbiology*, 16(3):255–261, 2013.

[32] Angela E Reeves, Mark J Koenigsknecht, Ingrid L Bergin, and Vincent B Young. Suppression of clostridium difficile in the gastrointestinal tracts of germfree mice inoculated with a murine isolate from the family lachnospiraceae. *Infection and immunity*, 80(11):3786–3794, 2012.

[33] A.M. Schubert, M.A. Rogers, C. Ring, J. Mogle, J.P. Petrosino, V.B. Young, D.M. Aronoff, and P.D. Schloss. Microbiome data distinguish patients with clostridium difficile infection and non-c. difficile-associated diarrhea from healthy controls. *mBio*, 5(3):e01021–14–e01021–14, 2014.

[34] C. Vincent, D.A. Stephens, V.G. Loo, T.J. Edens, M.A. Behr, K. Dewar, and A.R. Manges. Reductions in intestinal clostridiales precede the development of nosocomial clostridium difficile infection. *Microbiome*, 1(1):18, 2013.

[35] P. Singh, T.K. Teal, T.L. Marsh, J.M. Tiedje, R. Mosci, K. Jernigan, A. Zell, D.W. Newton, H. Salimnia, P. Lephart, D. Sundin, W. Khalife, R.A. Britton, J.T. Rudrik, and S.D. Manning. Intestinal microbial communities associated with acute enteric infections and disease recovery. *Microbiome*, 3(1), sep 2015.

[36] Julia MW Wong, Russell De Souza, Cyril WC Kendall, Azadeh Emam, and David JA Jenkins. Colonic health: fermentation and short chain fatty acids. *Journal of clinical gastroenterology*, 40(3):235–243, 2006.

[37] Gregory P Donaldson, S Melanie Lee, and Sarkis K Mazmanian. Gut biogeography of the bacterial microbiota. *Nature Reviews Microbiology*, 14(1):20–32, 2016.

[38] Joseph B. Kirsner and Roy G. Shorter. Recent developments in nonspecific inflammatory bowel disease. *New England Journal of Medicine*, 306(13):775–785, 1982.

[39] Marc Noguera-Julian, Muntsa Rocafort, Yolanda Guillén, Javier Rivera, Maria Casadellà, Piotr Nowak, Falk Hildebrand, Georg Zeller, Mariona Parera, Rocío Bellido, et al. Gut microbiota linked to sexual preference and hiv infection. *EBioMedicine*, 5:135–146, 2016.

[40] Catherine A Lozupone, Marcella Li, Thomas B Campbell, Sonia C Flores, Derek Linderman, Matthew J Gebert, Rob Knight, Andrew P Fontenot, and Brent E Palmer. Alterations in the gut microbiota associated with hiv-1 infection. *Cell host & microbe*, 14(3):329–339, 2013.

[41] D.M. Dinh, G.E. Volpe, C. Duffalo, S. Bhalchandra, A.K. Tai, A.V. Kane, C.A. Wanke, and H.D. Ward. Intestinal microbiota, microbial translocation, and systemic inflammation in chronic HIV infection. *Journal of Infectious Diseases*, 211(1):19–27, 2014.

[42] P.J. Turnbaugh, M. Hamady, T. Yatsunenko, B.L. Cantarel, A. Duncan, R.E. Ley, M.L. Sogin, W.J. Jones, B.A. Roe, J.P. Affourtit, M. Egholm, B. Henrissat, A.C. Heath, R. Knight, and J.I. Gordon. A core gut microbiome in obese and lean twins. *Nature*, 457(7228):480–484, 2008.

[43] J.K. Goodrich, J.L. Waters, A.C. Poole, J.L. Sutter, O. Koren, R. Blekhman, M. Beaumont, W. Van Treuren, R. Knight, J.T. Bell, T.D. Spector, A.G. Clark, and R.E. Ley. Human genetics shape the gut microbiome. *Cell*, 159(4):789–799, nov 2014.

[44] M.L. Zupancic, B.L. Cantarel, Z. Liu, E.F. Drabek, K.A. Ryan, S. Cirimotich, C. Jones, R. Knight, W.A. Walters, D. Knights, E.F. Mongodin, R.B. Horenstein, B.D. Mitchell, N. Steinle, S. Snitker, A.R. Shuldiner, and C.M. Fraser. Analysis of the gut microbiota in the old order Amish and its relation to the metabolic syndrome. *PloS one*, 7(8):e43052, 2012.

[45] M.C. Ross, D.M. Muzny, J.B. McCormick, R.A. Gibbs, S.P. Fisher-Hoch, and J.F. Petrosino. 16s gut community of the cameron county hispanic cohort. *Microbiome*, 3(1):7, 2015.

[46] Alexis Mosca, Marion Leclerc, and Jean P Hugot. Gut microbiota diversity and human diseases: should we reintroduce key predators in our ecosystem? *Frontiers in microbiology*, 7, 2016.

[47] Edoardo Pasolli, Duy Tin Truong, Faizan Malik, Levi Waldron, and Nicola Segata. Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. *PLoS Comput Biol*, 12(7):e1004977, 2016.

[48] David A Rasko, MJ Rosovitz, Garry SA Myers, Emmanuel F Mongodin, W Florian Fricke, Pawel Gajer, Jonathan Crabtree, Mohammed Sebaihia, Nicholas R Thomson, Roy Chaudhuri, et al. The pangenome structure of escherichia coli: comparative genomic analysis of e. coli commensal and pathogenic isolates. *Journal of bacteriology*, 190(20):6881–6893, 2008.

[49] Zi-Kai Wang and Yun-Sheng Yang. Upper gastrointestinal microbiota and digestive diseases. *World J Gastroenterol*, 19(10):1541–1550, 2013.

[50] Dwayne C Savage. Microbial ecology of the gastrointestinal tract. *Annual Reviews in Microbiology*, 31(1):107–133, 1977.

[51] Z. Zhang, H. Zhai, J. Geng, R. Yu, H. Ren, H. Fan, and P. Shi. Large-scale survey of gut microbiota associated with MHE via 16s rRNA-based pyrosequencing. *Am J Gastroenterol*, 108(10):1601–1611, jul 2013.

[52] J.U. Scher, A. Sczesnak, R.S. Longman, N. Segata, C. Ubeda, C. Bielski, T. Rostron, V. Cerundolo, E.G. Pamer, S.B. Abramson, C. Huttenhower, and D.R. Littman. Expansion of intestinal prevotella copri correlates with enhanced susceptibility to arthritis. *eLife*, 2, 2013.

[53] G. Falony, M. Joossens, S. Vieira-Silva, J. Wang, Y. Darzi, K. Faust, A. Kurilshikov, M. J. Bonder, M. Valles-Colomer, D. Vandeputte, R. Y. Tito, S. Chaffron, L. Rymenans, C. Verspecht, L. De Sutter, G. Lima-Mendez, K. Dhoe, K. Jonckheere, D. Homola, R. Garcia, E. F. Tigchelaar, L. Eeckhaudt, J. Fu, L. Henckaerts, A. Zhernakova, C. Wijmenga, and J. Raes. Population-level analysis of gut microbiome variation. *Science*, 352(6285):560–564, 2016.

[54] Lawrence A David, Arne C Materna, Jonathan Friedman, Maria I Campos-Baptista, Matthew C Blackburn, Allison Perrotta, Susan E Erdman, and Eric J Alm. Host lifestyle affects human microbiota on daily timescales. *Genome Biology*, 15(7):R89, 2014.

[55] Victoria Pascal, Marta Pozuelo, Natalia Borruel, Francesc Casellas, David Campos, Alba Santiago, Xavier Martinez, Encarna Varela, Guillaume Sarrabayrouse, Kathleen Machiels, Severine Vermeire, Harry Sokol, Francisco Guarner, and Chaysavanh Manichanh. A microbial signature for crohn's disease. *Gut*, 66(5):813–822, feb 2017.

[56] R. C. Edgar. Search and clustering orders of magnitude faster than blast. *Bioinformatics*, 26(19):2460–2461, 2010.

[57] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[58] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. [Online; accessed <today>].

[59] A.K. Alkanani, N. Hara, P.A. Gottlieb, D. Ir, C.E. Robertson, B.D. Wagner, D.N. Frank, and D. Zipris. Alterations in intestinal microbiota correlate with susceptibility to type 1 diabetes. *Diabetes*, 64(10):3510–3520, 2015.

[60] M.E. Mejía-León, J.F. Petrosino, N.J. Ajami, M.G. Domínguez-Bello, and A.M.C. de la Barca. Fecal microbiota imbalance in mexican children with type 1 diabetes. *Sci. Rep.*, 4, 2014.

[61] V.W. Wong, C. Tse, T.T. Lam, G.L. Wong, A.M. Chim, W.C. Chu, D.K. Yeung, P.T. Law, H. Kwan, J. Yu, J.J. Sung, and H.L. Chan. Molecular characterization of the fecal microbiota in patients with nonalcoholic steatohepatitis – a longitudinal study. *PLoS ONE*, 8(4):e62885, apr 2013.

Chapter 4

Correcting for batch effects in case-control microbiome studies

Sean M Gibbons, Claire Duvallet, and Eric J Alm

The figures appear within the main text of the chapter, and supplementary figures are at the end of the chapter.

Abstract

High-throughput data generation platforms, like mass-spectrometry, microarrays, and second-generation sequencing are susceptible to batch effects due to run-to-run variation in reagents, equipment, protocols, or personnel. Currently, batch correction methods are not commonly applied to microbiome sequencing datasets. In this paper, we compare different batch-correction methods applied to microbiome case-control studies. We introduce a model-free normalization procedure where features (i.e. bacterial taxa) in case samples are converted to percentiles of the equivalent features in control samples within a study prior to pooling data across studies. We look at how this percentile-normalization method compares to traditional meta-analysis methods for combining independent p-values and to limma and ComBat, widely used batch-correction models developed for RNA microarray data. Overall, we show that percentile-normalization is a simple, non-parametric approach for correcting batch effects and improving sensitivity in case-control meta-analyses.

Author summary

Batch effects are obstacles to comparing results across studies. Traditional meta-analysis techniques for combining p-values from independent studies, like Fisher's method, are effective but statistically conservative. If batch-effects can be corrected, then statistical tests can be performed on data pooled across studies, increasing sensitivity to detect differences between treatment groups. Here, we show how a simple, model-free approach corrects for batch effects in case-control microbiome datasets.

4.1 Introduction

Data generated by high throughput methods like mass-spectrometry, second-generation sequencing, or microarrays are sensitive to experimental and computational processing [1, 2]. This sensitivity gives rise to 'batch effects' between independent runs of an experiment. Even when different research groups adhere to the same methodologies,

these effects can arise due to slight differences in hardware, reagents, or personnel [3]. Thus, it is inappropriate to make direct, quantitative comparisons of uncorrected data across studies.

Several tools for reducing batch effects in RNA microarray data have been developed. For example, surrogate variable analysis (SVA) estimates a set of inferred variables (eigenvectors) that explain variance associated with putative batch effects [4]. These inferred variables are then incorporated into a linear model to correct downstream significance tests. The limma package employs a similar linear correction to account for batch effects prior to statistical analysis [5]. SVA and limma are part of a family of linear batch-correction methods that use different varieties of factor analysis, singular value decomposition, or regression [4, 5, 6, 7]. The most relied upon method to date [8], called ComBat, uses a Bayesian approach to estimate location and scale parameters for each feature within a batch [9]. All of these models are most effective when batch effects are not conflated with the true biological effects [1]. Furthermore, most batch correction methods make certain parametric assumptions.

Unfortunately, models that often work well for many types of 'omics data may not be appropriate for microbiome datasets. In microbiome studies, batch effects are often diffuse and conflated with biological signals [10, 11, 12]. The microbiome field has also struggled with finding appropriate parametric models for bacterial abundance distributions and for dealing with zeros. This is especially true for low-biomass samples in microbiome sequencing studies, like samples taken from the built environment [13], where populations are under-sampled, the biological signal is relatively weak, and batch effects can be quite large [14]. One way to get around this issue is to calculate statistics within a given batch, and then compare significant features across batches using classic meta-analysis techniques for combining p-values, like Fisher's and Stouffer's methods [15, 16]. These meta-analysis techniques are robust to batch effects across independent studies, but have less statistical power and ability to detect subtle differences than directly pooling data across studies.

Here, we describe a model-free data-normalization procedure for controlling batch effects in case-control microbiome studies that enables pooling data across studies.

Case-control studies include a built-in population of control samples (e.g. healthy subjects) that can be used to normalize the case samples (e.g. diseased subjects). For every feature (i.e. bacterial taxon), the case abundance distributions can be converted to percentiles of the equivalent control abundance distributions (Fig. 4-1). Study-specific batch effects present in the case samples will also be present in the control samples, and by converting the case data into percentiles of the control distribution these effects are mitigated. Upon conversion to percentiles of the within-study controls, percentile-normalized samples from multiple studies with similar case-control definitions can be more appropriately pooled for statistical testing (Fig. 4-1). We show that this approach effectively controls batch effects in microbiome case-control studies and we compare this method to pooling ComBat- or limma-corrected data, and to Fisher's and Stouffer's methods for combining independent p-values.

4.2 Methods

4.2.1 Datasets

We used a collection of case-control datasets obtained from the MicrobiomeHD database [17] to validate our batch-normalization method. We focused our analyses on studies spanning five diseases: colorectal cancer (CRC) [18, 12, 19, 20], Crohn's Disease (CD) [21, 22, 23, 24], Ulcerative Colitis (UC) [22, 23, 24], obesity (OB) [18, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34], and Clostridium difficile induced diarrhea (CDI) [32, 35]. For a subset of three CRC studies [18, 36, 19], we were able to obtain sequence data from the same region of the 16S gene (V4) so that these data could be processed together. The remaining MicrobiomeHD case-control datasets were previously processed using the same pipeline (see below), and then Operational Taxonomic Units (OTUs) were summarized at the genus level for comparison across studies.
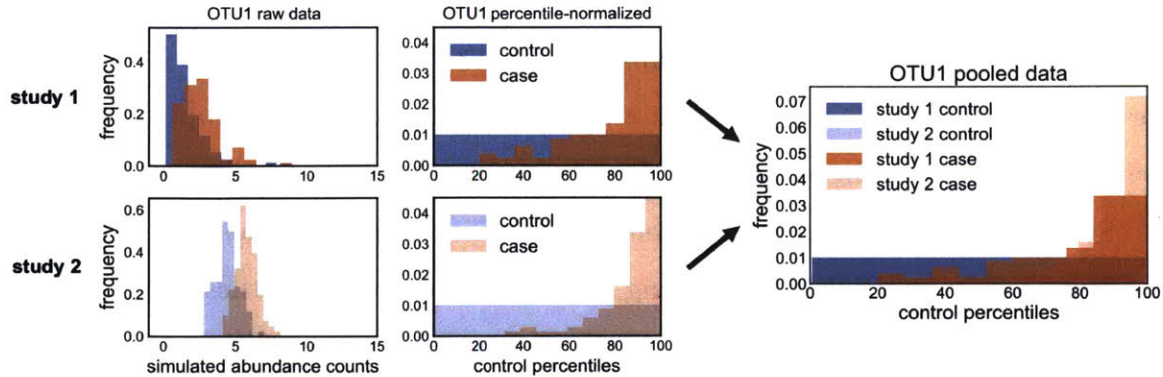
Figure 4-1: Percentile-normalization procedure converts case and control values into percentiles of the control distribution, which allows for pooling of normalized data across studies. Conceptual plot shows theoretical feature (OTU 1) abundance distributions for control samples and case samples from two independent studies. Converting a control distribution into percentiles of itself naturally gives rise to a uniform distribution (represented by flat blue distributions in central panels), while converting the case distribution into percentiles of the control distribution produces a non-uniform distribution when these two distributions differ (represented by skewed orange distributions in central panels). The right-most panel shows the result of pooling percentile distributions from study 1 and study 2. Percentile-normalization places data from separate studies onto a standardized axis that allows for cross-study comparison. Each simulated case and control distribution was produced by randomly sampling 100 times from a lognormal distribution. Study 1 control parameters: $\mu = 0.1$ and $\sigma = 0.7$. Study 1 case parameters: $\mu = 0.8$ and $\sigma = 0.5$. Study 2 control parameters: $\mu = 1.5$ and $\sigma = 0.2$. Study 2 case parameters: $\mu = 1.75$ and $\sigma = 0.13$.

4.2.2 Sequence Data Processing

To perform OTU-level analyses across the CRC studies, we downloaded the raw data from all of the MicrobiomeHD datasets that sequenced the V4 region of the 16S gene. We quality filtered and length trimmed each V4 dataset as described in Duvallet et al. (2017) and concatenated these raw, trimmed FASTQ files into one file. We removed any unique sequences that did not appear more than 20 times and clustered the remaining reads with USEARCH [37] at 97% similarity. We assigned these OTUs taxonomic identifiers using the RDP classifier [38] with a cutoff of 0.5. For genus-level analyses, OTU tables and metadata were acquired from the MicrobiomeHD database (`https://doi.org/10.5281/zenodo.569601`). Raw data were downloaded from the original studies and processed through our in-house 16S-processing pipeline (`https:`

`//github.com/thomasgurry/amplicon_sequencing_pipeline`) as described previously [17]. Each study's OTU table was converted to relative abundance by dividing each sample by its total number of reads and collapsed to genus level by summing all OTUs with the same genus-level annotation.

To plot data in ordination space, Bray-Curtis distances were calculated from relative abundance data using Scikit-learn (`sklearn.metrics.pairwise.pairwise_distances;` `metric='braycurtis'`) [39]. Non-metric multidimensional scaling (NMDS) coordinates were calculated for two axes based on Bray-Curtis distances using Scikit-learn (`sklearn.manifold.MDS; n_components=2, metric=False, max_iter=500, eps=1e-12,` `dissimilarity='precomputed'`).

4.2.3 Percentile Normalization

In this procedure, control feature distributions are percentile-normalized against themselves (resulting in a uniform distribution between 0 and 100) and case feature distributions are converted into percentiles of their equivalent control features. Treating our controls as null-hypotheses is motivated by the idea that healthy patients should be treated as similar across datasets, even though we understand that they will differ due to biological as well as technical batch effects. Relative abundance distributions were converted to percentiles using the SciPy v 0.19.0 [40] `stats.percentileofscore` method (`kind='mean'`). In order to avoid rank pile-ups due to the presence of many zeros, we replaced zeros with pseudo relative abundances drawn from a uniform distribution between 0.0 and 10^{-9} (i.e. a set of random values smaller than the lowest possible relative abundance in any dataset). Due to the zero-replacement step, p-values can shift slightly upon re-analysis with a different random draw, which can lead to the loss or gain of significance for features very near the significance threshold. Within each study, control distributions for each individual OTU or genus were converted into percentiles of themselves and case distributions were converted into percentiles of their corresponding control distribution. We have written a python script that performs percentile-normalization given an OTU table, a list of case sample IDs, and a list of control sample IDs as inputs (`https://github.com/seangibbons/percentile_`

normalization). A QIIME 2 (`https://qiime2.org`) plugin for running percentile-normalization is also available (`https://github.com/cduvallet/q2-perc-norm`).

4.2.4 ComBat

For each disease, we applied ComBat [8] to the case-control datasets analyzed in this study. Relative abundances (OTUs in the CRC analysis or OTUs collapsed to the genus level in the genus-level analysis) were log-transformed prior to running ComBat (default settings), adding a pseudo relative abundance of half the minimal frequency (across the entire feature table) to replace zeros. ComBat-corrected data were then transformed back from log-space (i.e. exponential transformation) prior to downstream analyses.

4.2.5 limma

In addition to ComBat, we applied a linear batch correction method from the limma package in R [5]. Relative abundances (zeros replaced with pseudo relative abundances equal to half the minimal frequency across the entire feature table) were log-transformed as described above and then a linear model was fit to subtract batch effects using the `removeBatchEffect` function (default settings). The limma-corrected data were then transformed back from log-space (i.e. exponential transformation) prior to downstream analysis.

4.2.6 Statistical Analysis

To calculate statistical significance, we restricted our statistical tests to OTUs/genera that occurred in at least one third of control or one third of case samples in order to reduce our multi-test correction penalty. We used the Wilcoxon rank-sum test, as implemented in SciPy v0.19.0 (`sicipy.stats.ranksums`) [40], to determine significant differences between independent groups of samples. Wilcoxon tests were run either within or across studies. In order to calculate statistics across studies, normalized case and control samples from multiple studies of the same disease were

combined together into the same OTU table. Hereafter, combining datasets is referred to as 'pooling.' P-values were multiple-test corrected using the Benjamini-Hochberg False Discovery Rate (FDR) procedure, as implemented in StatsModels v 0.8.0 (`statsmodels.sandbox.stats.multicomp.multipletests`) [41]. Differences in overall community structure were assessed using the Permutational Multivariate Analysis of Variance (PERMANOVA) test in R's vegan package [42] as implemented in scikit-bio (`skbio.stats.distance.permanova`). Fisher's and Stouffer's methods for combining p-values were performed using SciPy v0.19.0 (`cipy.stats.combine_pvalues`; `method='fisher'` or `method='stouffer'`). For Stouffer's method, weights for each study were calculated as the square root of the number of cases plus the number of controls. OTUs/genera with significant responses in opposing directions across studies were excluded from Fisher and Stouffer analyses.

4.2.7 *in silico* experiments

We ran an *in silico* titration experiment using the OTU-level data to simulate pooling of control samples from different datasets before calculating significant differences. Healthy samples from one study were mixed with healthy samples from another study at different proportions prior to calculating significant differences in OTU frequencies between cases and controls. Case and control groups were subsampled to 40 samples each. Control samples were substituted by randomly selected samples from another study along a fractional gradient (0-100% control samples from another study). We calculated significant differences between case and control groups using the Wilcoxon rank-sum test and applied an FDR correction. OTUs with q-values ≤ 0.05 were considered significant. The titration experiment was rerun 20 times, and the results were averaged.

Similar to the titration experiment, we ran an OTU-level analysis of how batch-correction methods might impact false-positive rates by randomly selecting 40 control samples from the Baxter et al. (2016) study as artificial 'controls' and 40 control samples from the Zeller et al. (2014) study as artificial 'cases' (across 20 iterations) for each data type (i.e. raw, percentile-normalized, limma-corrected, and ComBat-

corrected). We then calculated significant differences between these artificial 'case' and 'control' groups as outlined above to generate p-values for each OTU.

4.3 Results

4.3.1 Batch effects at OTU-level resolution

To minimize possible biases across data sets, we identified three colorectal cancer (CRC) studies that sequenced the same region of the 16S gene (V4). We reprocessed the raw sequence data from each study in the same quality filtering and OTU picking pipeline to obtain bioinformatically-standardized results. OTUs that occurred in at least one third of case or one third of control samples (i.e. within individual studies) were retained for all downstream statistical analyses. Despite standardizing the bioinformatic processing of these data, we saw significant batch effects in healthy patients across studies (PERMANOVA p < 0.001; Fig. 4-2). The similarity between samples from the Baxter et al. (2016) and Zackular et al. (2014) studies is due to the fact that they were sourced from the same patient cohort (although samples were processed separately), making this comparison a good pseudo-negative control for batch effects [18, 19]. There was an apparent reduction in the batch effect after applying ComBat, although differences between batches remained weakly significant (PERMANOVA p = 0.008, Fig. 4-2) [8]. Due to the non-independence between the Baxter and Zackular patient cohorts, we removed the smaller of the two studies (Zackular) from all downstream analyses. Out of a total of 1,021 OTUs that passed our abundance filter, 681 differed significantly in uncorrected relative abundance between the Baxter and Zeller healthy controls (FDR q ≤ 0.05).

We ran an *in silico* titration experiment to simulate pooling of control samples from different datasets before calculating significant differences. Healthy samples from one study were mixed with healthy samples from another study at different proportions prior to calculating significant differences in OTU frequencies between cases and controls (see conceptual outline in Fig. 4-3). For non-normalized data, the
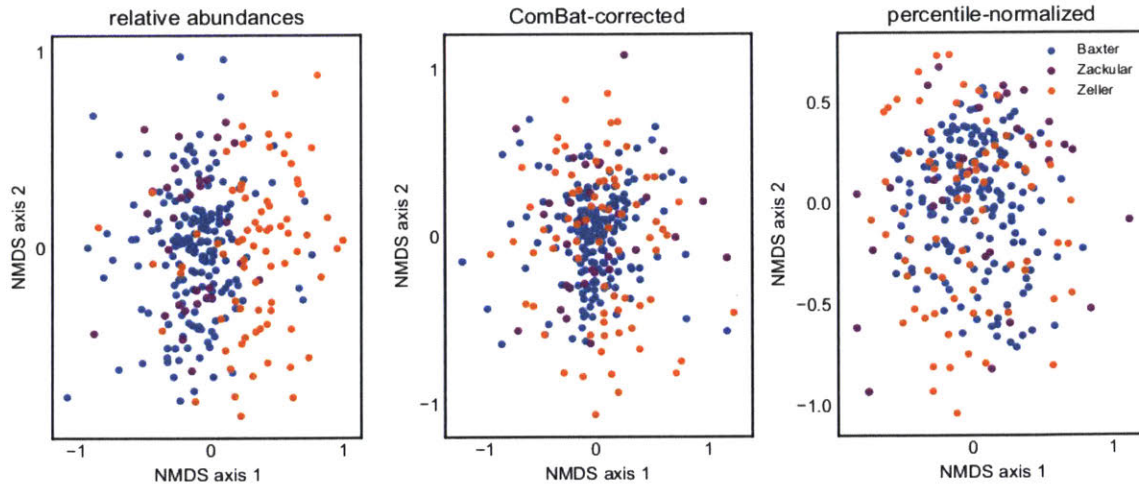
Figure 4-2: Batch effects between healthy controls from different studies can be reduced by ComBat and percentile-normalization. Non-metric multidimensional scaling (NMDS) plot showing the distribution of healthy controls from three colorectal cancer studies in ordination space (Bray-Curtis distances of relative abundance OTU-level data). Despite standardized bioinformatic processing, healthy patients differed significantly in their gut microbiomes across studies (PERMANOVA $p < 0.001$; batch accounts for 6.342% of the total variance). Studies were still significantly different after applying ComBat, an established batch-correction method (PERMANOVA $p < 0.01$). However, percentile-normalization did a better job of stabilizing the variance across studies and removed any apparent batch effect (PERMANOVA $p > 0.5$).

number of significant OTUs greatly increased due to batch effects as more control samples were substituted in from another study. This result highlights the danger of pooling raw data across batches. ComBat- and limma-corrected data performed better than uncorrected data, but still showed many spurious results as the proportion of control samples from another study increased (Fig. 4-3). Percentile-normalization showed no increase in spurious results over the titration gradient (Fig. 4-3). Although we do see several significant CRC-associated OTUs in the full dataset (see below), these were not detected in the titration experiment due to the loss of statistical power when reducing case and control groups to only 40 samples each.

We ran a second *in silico* experiment to determine whether the false-positive rate was impacted by our different batch-correction methods. We randomly selected 40 samples from the Baxter healthy controls as artificial 'controls' and 40 samples from
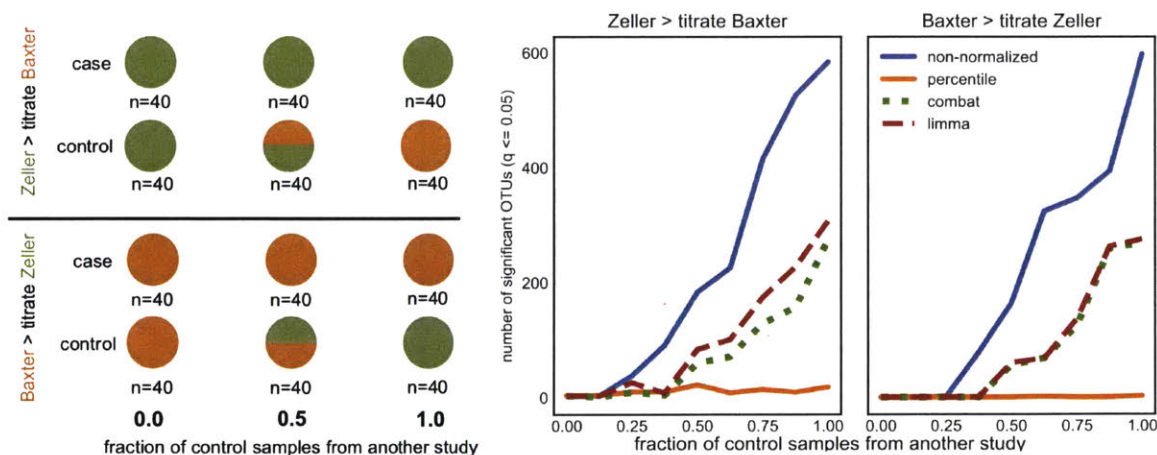
Figure 4-3: Pooling non-normalized samples from different studies can give rise to many spurious associations. The control group from one study is gradually substituted with randomly chosen control samples from another study (non-normalized, percentile-normalized, limma-corrected, and ComBat-corrected), keeping the total number of case and control samples fixed at n = 40 (see conceptual illustration on the left). Mixing in non-normalized control samples from another study gave rise to spurious results due to batch effects (blue lines). ComBat- and limma-corrected data showed fewer spurious associations (green and red lines). Percentile-normalization showed no increase in spurious results along the titration gradient (orange lines).

the Zeller healthy controls as artificial 'cases' for each data type (i.e. non-normalized, percentile-normalized, limma-corrected, and ComBat-corrected) and calculated significant OTU-level differences between these groups. We repeated this process twenty times to generate a set of p-value distributions. We found that the fraction of p-values ≤ 0.05 can be as high as 70% for the non-normalized data (Fig. 4-4). This result matches with our finding that 681 out of the 1,021 OTUs in this dataset differed significantly across Baxter and Zeller controls ($q \leq 0.05$). Each normalization technique drastically reduced the number of false positives, but percentile-normalization gave the best results (Fig. 4-4). When low-abundance OTUs were included in the analysis, ComBat and limma showed highly skewed p-value distributions, giving rise to a larger number of false positives than the non-normalized data (Fig. 4-8).

We next assessed the performance each cross-study analysis method by comparing OTU-level results across two independent CRC datasets. In the Baxter study, there were 172 healthy (control) samples and 120 CRC (case) samples, with 14
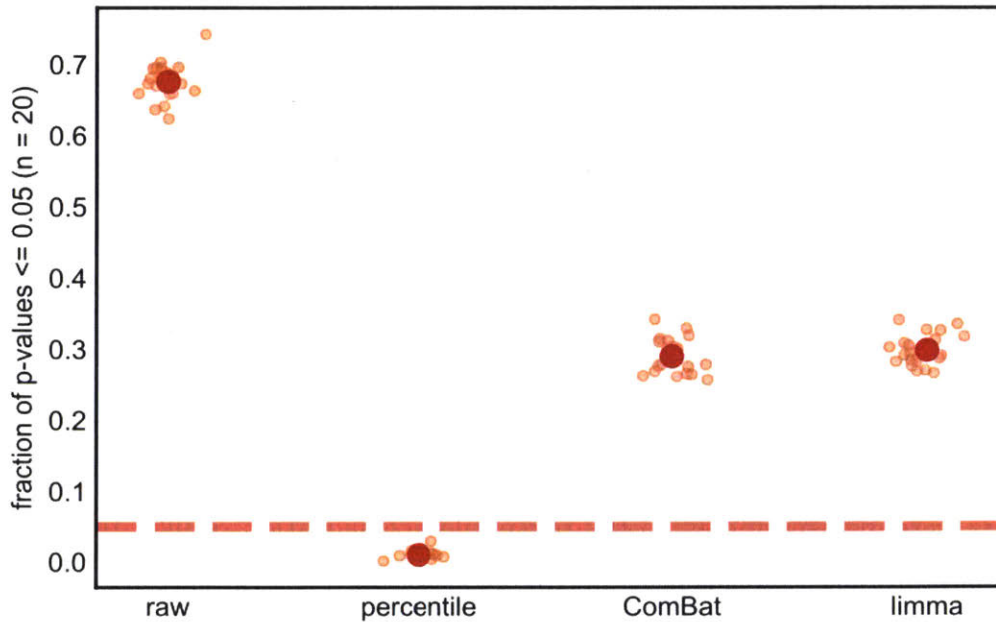
90

Figure 4-4: False positive rates are reduced by batch-correction methods. Random sets of 40 Baxter controls and random sets of 40 Zeller controls were selected for null case-control comparisons (20 iterations). Smaller points show the fraction of p-values ≤ 0.05 within a given iteration, while larger dots show the average value across all 20 iterations. Within each category, smaller points are randomly jittered along the x-axis for better visualization. The fraction of p-values ≤ 0.05 is highly inflated for non-normalized data (red dashed line shows the null-expectation for p-values). Only abundant OTUs (detected in at least a third of case or control samples) were included in this analysis.

OTUs (from Fusobacterium, Coprococcus, Butyricicoccus, Gemmiger, Faecalibacterium, Roseburia, Parvimonas, Haemophilus, Porphyromonas, Peptostreptococcus, Streptophyta, Bacteroides and Clostridium XIVa genera) showing significant differences in abundance between cases and controls (FDR q ≤ 0.05). For Zeller, there were 71 control and 40 case samples, with 18 OTUs (from Butyricicoccus, Butyricimonas, Fusobacterium, Closridium XIVa, Streptococcus, Parabacteroides, Alistipes, Anaerostipes, Parvimonas, Peptostreptococcus, Blautia, Dialister, and Bacteroides genera) that differed significantly across cases and controls (FDR q ≤ 0.05).

In the absence of batch effects, pooling data across datasets of the same dis-

ease should increase sensitivity to detect significant cross-study associations. We pooled percentile-normalized, limma-corrected, and ComBat-corrected data, respectively, across Baxter and Zeller studies to look for OTUs that differed significantly across cases and controls. These pooled results were then compared to classic methods for combining p-values from each dataset's individual results (above). For the percentile-normalized data, we found 39 OTUs that differed significantly between cases and controls (FDR q $\leq$ 0.05), 21 of which overlapped with the within-study results. The pooled limma-corrected and ComBat-corrected data resulted in 37 and 36 significant OTUs, respectively. 35 of the OTUs identified as significant by ComBat were also significant in the limma results. 30 of the limma results and 29 of the Com-Bat results were also significant in the percentile-normalization results, respectively. Fisher's method identified seven significant OTUs from Clostridium XIVa, Streptococcus, Fusobacterium, Parvimonas, Peptostreptococcus, and Anaerostipes genera, which were also found in the percentile-normalized results. Stouffer's method identified the same seven OTUs found using Fisher's method. Overall, the pooling methods improve statistical power to detect significant OTUs over traditional meta-analysis methods. For example, particular OTUs from Desulfovibrio and Parabacteroides genera were identified as significantly enriched in CRC patients in the pooled results (ComBat, limma, and percentile-normalized), but not in the within-study results or in the Fisher and Stouffer results. Pooled analysis of percentile-normalized data also identified on Enterobacter OTU enriched in cancer patients, two OTUs from the Lachnospiraceae family that were enriched in controls and one Lachnospiraceae that was enriched in cases, which were missed by the within-study analyses. In all, 18 OTUs were identified in the pooled, percentile-normalized results that were missed by the within-study analyses (Fig. 4-5). These additional taxonomic associations (e.g. Desulfovibrio, Costridium XIVa, and Lachnospiraceae) are consistent with prior meta-analyses of CRC microbiome studies [17, 43]. It is important to visualize the data being fed into statistical tests to determine whether significant associations are being driven by outlier studies or by other artifacts. The associations identified in Figure 4-5 appear to be biologically meaningful due to the overall consistency of the

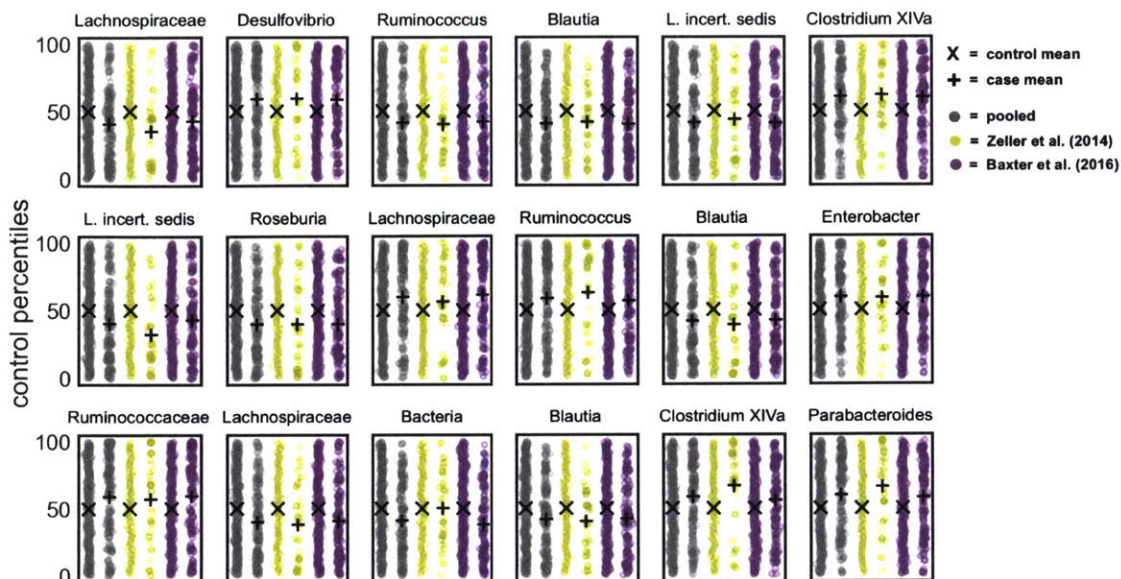effect directions across studies.

Figure 4-5: OTUs significant across CRC studies, but not within a given study. Pooling data provides greater statistical power to detect subtle, yet consistent differences in OTU abundances across sample groups. 18 OTUs are labeled by their most resolved taxonomic annotation. Each OTU in this plot was not found to be significant within either Baxter or Zeller studies, but became significant after pooling the percentile-normalized datasets ($q \leq 0.05$).

4.3.2 Batch effects at genus-level resolution across multiple diseases

In order to assess the performance of different meta-analysis techniques across a larger set of studies and diseases, we summarized OTU abundances at the genus level for five diseases across 18 studies – Clostridium difficile induced diarrhea (CDI), Crohn's disease (CD), ulcerative colitis (UC), obesity (OB), and CRC. There were a total of 306 unique genera detected across studies. There were two CDI case-control studies: Schubert et al. (2014) had 154 control and 93 case samples [32]; Vincent et al. (2013) had 25 control and 25 case samples [35]. There were four inflammatory bowel disease (IBD) studies that included CD patients and three that also included UC patients: Papa et al. (2012) had 24 non-IBD control samples, 23 CD samples, and 43 UC

93

samples [22]; Morgan et al. (2012) had 18 control, 61 CD and 47 UC samples [23]; Willing et al. (2010) had 35 control, 16 UC and 29 CD samples [24]; Gevers et al. (2014) had 16 non-IBD control and 146 CD samples, with no UC samples [21]. There were eleven studies with lean and obese (OB) cohorts: Turnbaugh et al. (2009) had 33 controls and 102 cases [25]; Goodrich et al. (2014) had 428 controls and 185 cases [29]; Escobar et al. (2014) had 10 controls and 10 cases [26]; Zhu et al. (2014) had 16 controls and 25 cases [44]; Jumpertz et al. (2011) had 12 controls and 9 cases [33]; Ross et al. (2015) had 26 controls and 37 cases [28]; Zupancic et al. (2012) had 96 controls and 101 cases [27]; Baxter et al. (2016) had 125 controls and 47 OB cases [18]; Schubert et al. (2014) had 68 controls and 34 OB cases [32]; Wu et al. (2011) had 59 controls and 9 cases [31]; and Zeevi et al. (2015) had 567 controls and 151 cases [30]. There were four independent CRC studies, including the Baxter and Zeller studies listed in the OTU-level analysis (see above for sample sizes). The remaining two CRC studies are Wang et al. (2012), which had 54 control and 44 case samples [45], and Chen et al. (2012), which had 22 controls and 21 cases [20].

Most genera that differed significantly within a given study were not significant in other studies of the same disease. For example, of the 36 unique genera that showed significant differences within any given OB study, none were found to be significant in all studies (I = 0; Table 4.1). Indeed, there were no genera that were significant across all studies in the majority of diseases studied (I = 0; Table 4.1). CDI only had two studies, and of the 38 significant results, only six were shared across both datasets. Overall, few genera were significant within two or more studies (2N $\leq$ 6; Table 4.1). The number of genera that differed significantly across pooled cases and controls changed depending on how the data were batch-corrected (Table 4.1). For every disease, percentile-normalization yielded the largest number of significant genera when compared to other methods. Overall, ComBat- and limma-corrections resulted in many fewer significant genera, especially for CD, UC, and CRC (Table 4.1). Half of the IBD (CD and UC) studies included non-IBD patients with inflammatory symptoms as controls rather than clinically healthy patients. These biologically relevant differences in inflammatory symptoms between control cohorts were con-

flated with batches and were likely smoothed out by ComBat and limma corrections. Fisher's and Stouffer's methods consistently identified fewer significant associations than percentile-normalization (Table 4.1). Pooling data prior to running a statistical test is a more sensitive technique than combining independently calculated p-values [46]. Thus, percentile-normalization increases the statistical power to detect differences across studies while controlling for false positives and batch effects.

disease	method	pooled	within
CDI (N=2)	percentile	37	U = 38 / 2N = 6
	Fisher	12	
	Stouffer	12	
	ComBat	36	
	limma	36	
CD (N=4)	percentile	19	U = 13 / I = 0 / 2N = 2
	Fisher	6	
	Stouffer	6	
	ComBat	2	
	limma	1	
UC (N=3)	percentile	10	U = 17 / I = 0 / 2N = 1
	Fisher	4	
	Stouffer	4	
	ComBat	5	
	limma	5	
CRC (N=4)	percentile	12	U = 20 / I = 0 / 2N = 3
	Fisher	9	
	Stouffer	7	
	ComBat	5	
	limma	5	
OB (N=11)	percentile	18	U = 36 / I = 0 / 2N = 6
	Fisher	4	
	Stouffer	6	
	ComBat	13	
	limma	15	

Table 4.1: Normalization methods impact the number of significant genus-level associations between cases and controls across multiple diseases. Numbers of genera that differ significantly between cases and controls for five diseases. In the 'disease' column, CDI = Clostridium difficile induced diarrhea, CD = Crohn's Disease, UC = Ulcerative Colitis, CRC = Colorectal Cancer, and OB = obesity. 'N =' shows the number of studies included in each meta-analysis. The method column indicates how the data were processed prior to running significance tests (percentile-normalized, Fisher's method for combining p-values, Stouffer's method for combining p-values, ComBat-corrected, or limma-corrected). The significance threshold used was $q \leq 0.05$ (FDR). The 'pooled' column shows the total number of genera that were found to be significantly different across pooled studies for a given disease. The 'within' column shows the total number of unique (non-redundant) genera that were identified as significantly different within each study (U = union), the number of genera that were significant in all individual studies (I = intersection), and the number of significant genera that were consistently significant in at least two studies (2N; 2N == I for CDI).

To better assess how percentile normalization impacted the pooled results, we looked at genera that were significant within a single-study but not across studies after pooling. There were 12 genera that were significant within a subset of CRC studies, but not after pooling (Fig. 4-6). Gemmiger, Bacteroides, and Roseburia showed variable responses across studies, sometimes enriched in controls and other times enriched in cases. The remaining genera showed weak associations within one or two studies, but did not differ significantly across studies (i.e. $q \leq 0.05$). These genera that show weak or inconsistent responses across batches may not be reliable disease biomarkers. However, by including larger numbers of CRC studies in future meta-analyses it is likely that some of these genera could pass the significance threshold. Two genera - Lactobacillus and Desulfovibrio - were not significantly different between cases and controls within an individual study, but became significant after pooling (Fig. 4-7). These genera showed weak, but largely consistent enrichment in cancer patients and demonstrate the utility of pooling datasets to detect subtle differences.
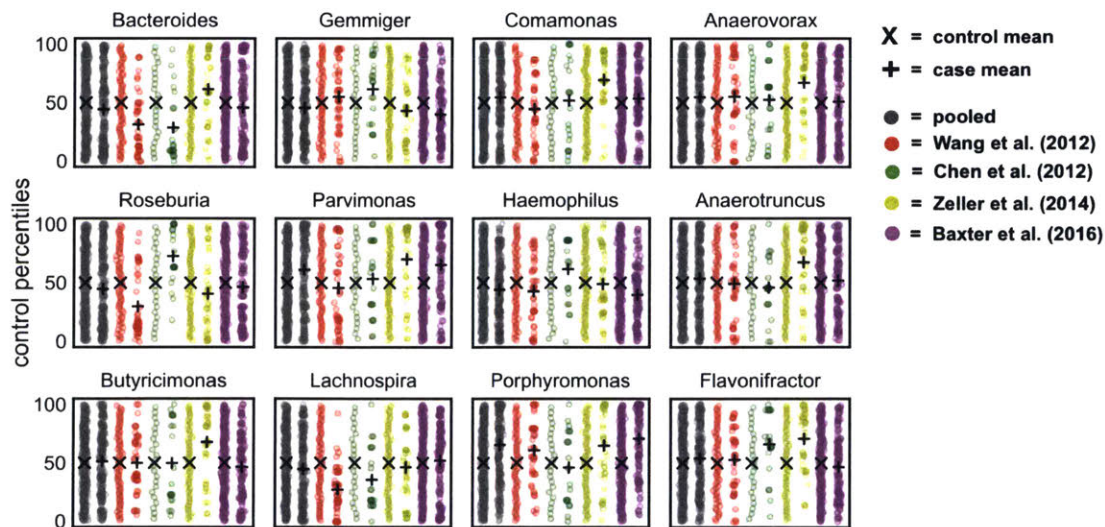

Figure 4-6: Genera that show a significant difference between CRC cases and controls within a given study, but not after pooling. 12 genera showed significant differences between cases and controls within a study ($q \leq 0.05$), but not after pooling across CRC studies.

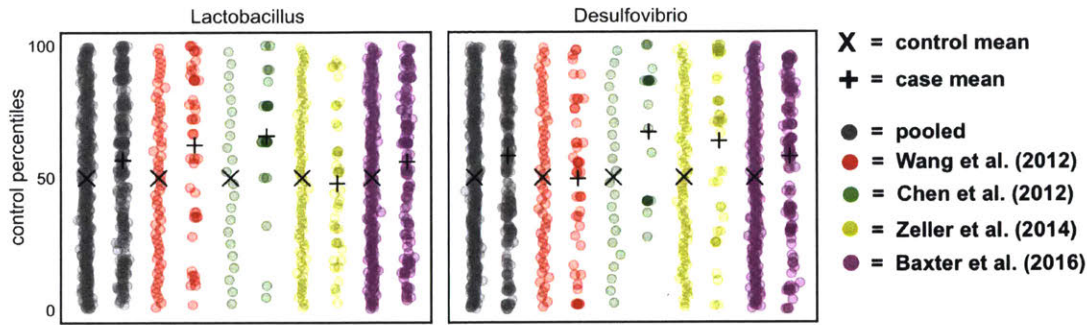While prior work has suggested that there may not be consistent associations

Figure 4-7: Genera that do not show a significant difference between CRC cases and controls within a given study, but do after pooling. Two genera did not show significant differences between cases and controls within a study, but became significant after pooling across CRC studies ($q \leq 0.05$).

between the gut microbiome and obesity [34], we observed six genera in a recent meta-analysis that differed significantly across two or more (out of five) independent obesity studies [17]. Of these six genera, four (Roseburia, Clostridium IV, Oscillobacter, and Pseudoflavonifractor) were also found to be significant in the pooled, percentile-normalized results (Fig. 4-9). The two remaining genera not found to be significant in the percentile-normalized analysis (Mogibacterium and Anaerovorax) showed highly irregular responses across the 11 obesity studies analyzed in this study (Fig. 4-10). Despite the irregular behavior of these genera, Fisher's and Stouffer's methods both identified Mogibacterium as significantly associated with obesity.

4.4 Discussion

Batch effects are unavoidable when working with high-throughput data generation platforms. The RNA microarray community has been proactive in the development of tools for dealing with these effects [1, 8]. However, these tools are not as effective when batch effects are confounded with biological signals or when parametric assumptions do not apply, which is often the case in microbiome case-control studies. Therefore, model-free methods are needed for correcting batch effects across microbiome datasets. Fortunately, case-control studies can be internally normalized by

their own control samples. Any study-specific batch effects in the case samples will be present in the control samples, and by converting the case data into percentiles of the control distribution these effects are attenuated without making parametric assumptions.

Relative abundance, limma-corrected, and ComBat-corrected data – but not percentile-normalized data – quickly yielded a large number of spurious results when cases from one study were tested against controls from another (Fig. 4-3). Additionally, when control populations from different batches were compared to one another, non-normalized data yielded a much larger number of false positives than batch-corrected data (Fig. 4-4). Our percentile-normalization approach was much more effective than limma and ComBat in controlling false positives (Fig. 4-4), especially in the presence of low-abundance taxa (Fig. 4-8).

Because pooling datasets increases statistical power, it is tempting to pool these data even in the absence of suitable batch-correction methods. Consequently, pooling non-normalized data from different batches has been common practice in the microbiome field [26, 28, 34, 47, 48, 49, 50, 51, 52, 53, 54, 55]. In this paper, we demonstrate why this practice is highly inadvisable. Pooling batch-corrected data from multiple studies allowed us to detect significant differences that were not found within a given study (Figs. 4-5 and 4-7), while removing associations that were weak or inconsistent across studies (Fig. 4-6 and 4-10). Percentile-normalized results often identified significant differences between cases and controls that were missed by other normalization methods (Table 4.1). For CDI, percentile-normalized results identified about the same number of significant hits as the other batch-correction methods (Table 4.1), which was likely due to the fact that the biological signal associated with diarrhea is very strong [17]. In cases where the biological signal is strong, results should be robust to the types of analyses employed. For UC and CD studies (IBD), percentile-normalization identified several significant genera that limma and ComBat did not (Table 4.1). The reduced number of significant hits from limma- and ComBat-corrected data for IBD was likely due to heterogeneous control cohorts across these studies (i.e. healthy patients vs. non-IBD patients), which likely

smoothed-out inflammation-associated signals. This result highlights the importance of having consistent definitions for case and control cohorts across studies.

We compared percentile-normalization and pooling to Fisher's and Stouffer's methods for combining independent p-values. Stouffer's method is similar to Fisher's, but includes weights for each p-value based on the number of samples in a study. Percentile-normalization consistently identified a larger number of significant hits than Stouffer's and Fisher's methods, confirming that pooling data increases sensitivity (i.e. reducing putative false negatives). Methods for combining p-values from independent studies are quite robust and should probably be considered as a safe alternative to pooling (i.e. lower chance of false positives). However, in the case of our obesity analysis, Fisher's and Stouffer's methods identified Mogibacterium as significant despite its apparent inconsistency across studies (Fig. 4-10).

In conclusion, we present a robust, model-free procedure for transforming each feature in a microbiome case-control dataset into percentiles of its control distribution (Fig. 4-1). The main conditions for applying this method are that 1) each batch must have a sizeable number of control samples (i.e. the density of the control distribution limits the resolution of the percentile-transformation of the case samples), and 2) case and control populations should be consistently defined across batches (i.e. same definition of 'healthy' or 'diseased' groups). Given these caveats, percentile-normalized features can be pooled across studies for univariate statistical testing (whichever test a researcher prefers – ideally non-parametric), alleviating the batch effect problem. This model-free procedure could also be applied to other types of 'omics datasets with consistently defined internal controls. We find that this procedure allows us to identify differences between cases and controls that are often missed by more conservative meta-analysis techniques. Methods developed for batch-correction in microarray data, like limma and ComBat, can partially reduce batch effects in microbiome studies (Figs. 4-2–4-4), but appear to obscure real patterns if batch effects are not independent of biological signals or if the parametric assumptions of these models are not valid. We suggest that methods like limma and ComBat are useful for studies lacking case and control groups. However, when studies have consistently defined

internal controls, percentile-normalization should be the preferred batch correction approach. Future work should focus on developing parametric models specifically for batch correction in microbiome datasets, which could further improve sensitivity to detect subtle biological differences across studies.

4.5 Acknowledgements
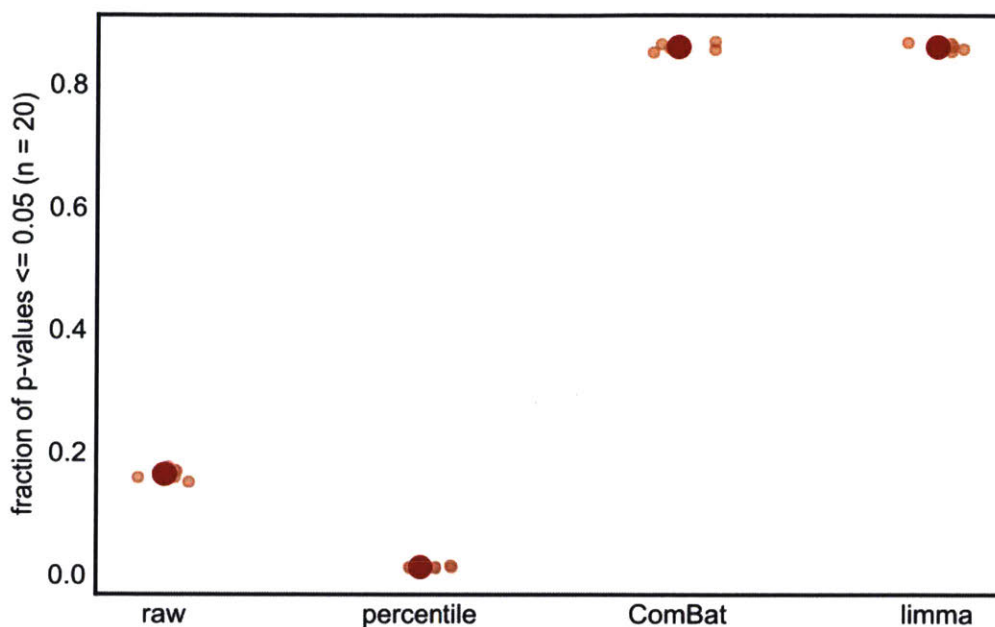
4.6 Supplementary Figures

Figure 4-8: False positive rates are reduced by percentile-normalization, but not by ComBat or limma, in the presence of low-abundance OTUs. Random sets of 40 Baxter controls and random sets of 40 Zeller controls were selected for null case-control comparisons (20 iterations). Smaller points show the fraction of p-values $\leq$ 0.05 within a given iteration, while larger dots show the average value across all 20 iterations. Within each category, smaller points are randomly jittered along the x-axis for better visualization. The fraction of p-values $\leq$ 0.05 is highly inflated for all methods except percentile-normalization (red dashed line shows the null-expectation for p-values). All OTUs were included in this analysis (i.e. no abundance filter prior to running tests). ComBat and limma show highly skewed p-value distributions when including low-abundance OTUs.
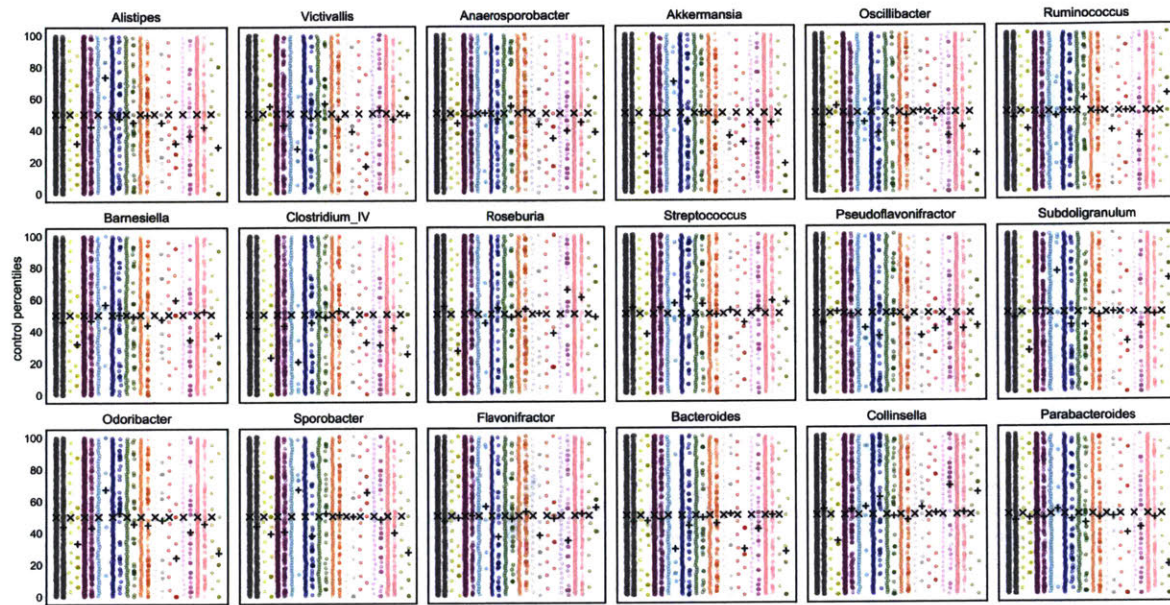
Figure 4-9: 18 genera significantly different (q ≤ 0.05) between obese and healthy patients based on pooled percentile-normalized analysis (from Table 4.1). Gray points show data pooled from all 11 obesity studies. Other colors show individual obesity studies. Studies listed in order from left to right: Zhu et al. (2014), Zeevi et al. (2015), Wu et al. (2011), Baxter et al. (2016), Schubert et al. (2014), Zupancic et al. (2012), Ross et al. (2015), Jumpertz et al. (2011), Turnbaugh et al. (2009), Goodrich et al. (2014), Escobar et al. (2014). X symbols indicate the average control percentiles, while + symbols indicate the average case percentiles.
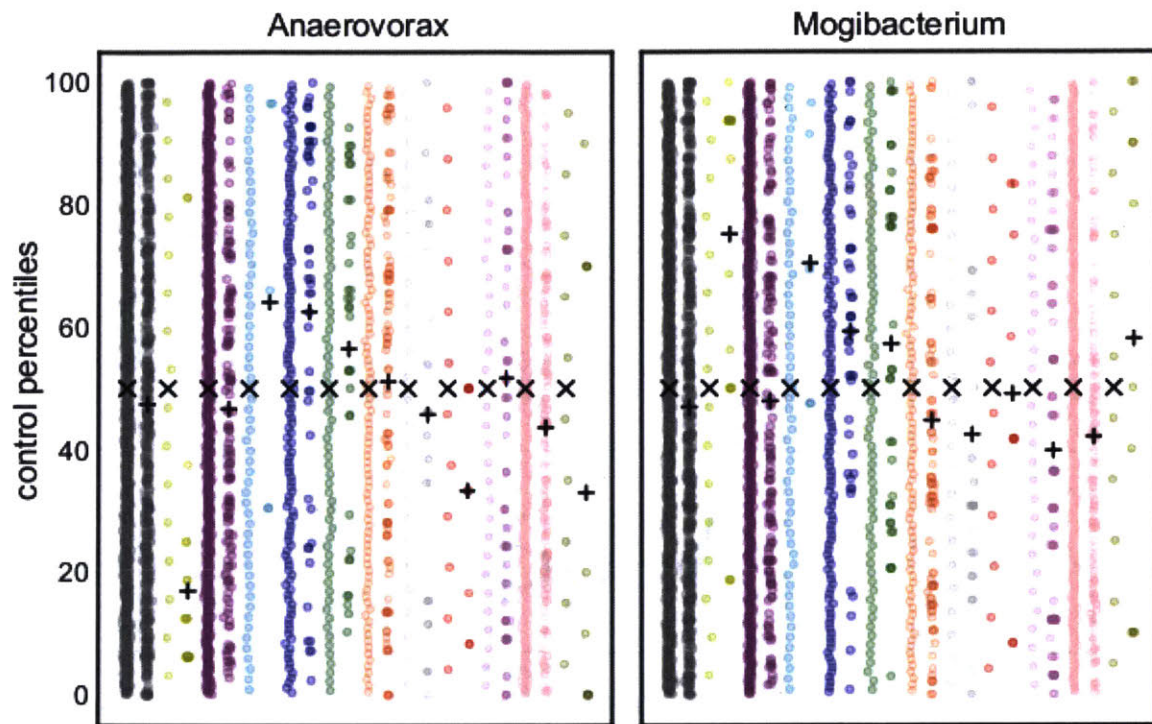
Figure 4-10: Two genera that were significant (q ≤ 0.05) in at least two obesity studies in Duvallet et al. (2017), but were not significant in the pooled, percentile-normalized analysis. Gray points show data pooled from all 11 obesity studies. Other colors show individual obesity studies. Studies listed in order from left to right: Zhu et al. (2014), Zeevi et al. (2015), Wu et al. (2011), Baxter et al. (2016), Schubert et al. (2014), Zupancic et al. (2012), Ross et al. (2015), Jumpertz et al. (2011), Turnbaugh et al. (2009), Goodrich et al. (2014), Escobar et al. (2014). X symbols indicate the average control percentiles, while + symbols indicate the average case percentiles.

Bibliography

[1] Jeffrey T. Leek, Robert B. Scharpf, Héctor Corrada Bravo, David Simcha, Benjamin Langmead, W. Evan Johnson, Donald Geman, Keith Baggerly, and Rafael A. Irizarry. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, 11(10):733–739, sep 2010. doi: 10.1038/nrg2825. URL https://doi.org/10.1038/nrg2825.

[2] Wilson Wen Bin Goh, Wei Wang, and Limsoon Wong. Why batch effects matter in omics data, and how to avoid them. *Trends in Biotechnology*, 35(6):498–507, jun 2017. doi: 10.1016/j.tibtech.2017.02.012. URL https://doi.org/10.1016/j.tibtech.2017.02.012.

[3] Patrick D. Schloss, Dirk Gevers, and Sarah L. Westcott. Reducing the effects of PCR amplification and sequencing artifacts on 16s rRNA-based studies. *PLoS ONE*, 6(12):e27310, dec 2011. doi: 10.1371/journal.pone.0027310. URL https://doi.org/10.1371/journal.pone.0027310.

[4] Jeffrey T. Leek and John D. Storey. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genetics*, 3(9):e161, 2007. doi: 10.1371/journal.pgen.0030161. URL https://doi.org/10.1371/journal.pgen.0030161.

[5] Matthew E. Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):e47–e47, jan 2015. doi: 10.1093/nar/gkv007. URL https://doi.org/10.1093/nar/gkv007.

[6] O. Alter, P. O. Brown, and D. Botstein. Singular value decomposition for genome-wide expression data processing and modeling. *Proceedings of the National Academy of Sciences*, 97(18):10101–10106, aug 2000. doi: 10.1073/pnas.97.18.10101. URL https://doi.org/10.1073/pnas.97.18.10101.

[7] M. Benito, J. Parker, Q. Du, J. Wu, D. Xiang, C. M. Perou, and J. S. Marron. Adjustment of systematic microarray data biases. *Bioinformatics*, 20(1):105–114, dec 2003. doi: 10.1093/bioinformatics/btg385. URL https://doi.org/10.1093/bioinformatics/btg385.

[8] Chao Chen, Kay Grennan, Judith Badner, Dandan Zhang, Elliot Gershon, Li Jin, and Chunyu Liu. Removing batch effects in analysis of expression microarray data: An evaluation of six batch adjustment methods. *PLoS ONE*, 6(2):e17238, feb 2011. doi: 10.1371/journal.pone.0017238. URL https://doi.org/10.1371/journal.pone.0017238.

[9] W. Evan Johnson, Cheng Li, and Ariel Rabinovic. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics*, 8(1): 118–127, apr 2006. doi: 10.1093/biostatistics/kxj037. URL https://doi.org/10.1093/biostatistics/kxj037.

[10] Sophie Weiss, Amnon Amir, Embriette R Hyde, Jessica L Metcalf, Se Jin Song, and Rob Knight. Tracking down the sources of experimental contamination in microbiome studies. *Genome Biology*, 15(12), dec 2014. doi: 10.1186/s13059-014-0564-2. URL https://doi.org/10.1186/s13059-014-0564-2.

[11] Hua Shen, Snezna Rogelj, and Thomas L. Kieft. Sensitive, real-time PCR detects low-levels of contamination by legionella pneumophila in commercial reagents. *Molecular and Cellular Probes*, 20(3-4):147–153, jun 2006. doi: 10.1016/j.mcp.2005.09.007. URL https://doi.org/10.1016/j.mcp.2005.09.007.

[12] Nhu H. Nguyen, Dylan Smith, Kabir Peay, and Peter Kennedy. Parsing ecological signal from noise in next generation amplicon sequencing. *New Phytologist*, 205 (4):1389–1393, jul 2014. doi: 10.1111/nph.12923. URL https://doi.org/10.1111/nph.12923.

[13] Sean M. Gibbons. The built environment is a microbial wasteland. *mSystems*, 1(2), apr 2016. doi: 10.1128/msystems.00033-16. URL https://doi.org/10.1128/msystems.00033-16.

[14] John Chase, Jennifer Fouquier, Mahnaz Zare, Derek L. Sonderegger, Rob Knight, Scott T. Kelley, Jeffrey Siegel, and J. Gregory Caporaso. Geography and location are the primary drivers of office microbiome composition. *mSystems*, 1(2), apr 2016. doi: 10.1128/msystems.00022-16. URL https://doi.org/10.1128/msystems.00022-16.

[15] Ronald A. Fisher. *Statistical methods for research workers*. Edinburgh: Oliver and Boyd, 1925.

[16] S.A. Stouffer, E.A. Suchman, L.C. Devinney, S.A. Star, and R.M. Williams. *The American soldier: Adjustment during army life. (Studies in social psychology in World War II)*. Oxford, England: Princeton Univ. Press, 1949.

[17] Claire Duvallet, Sean M. Gibbons, Thomas Gurry, Rafael A. Irizarry, and Eric J. Alm. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nature Communications*, 8(1), dec 2017. doi: 10.1038/s41467-017-01973-8. URL https://doi.org/10.1038/s41467-017-01973-8.

[18] Nielson T. Baxter, Mack T. Ruffin, Mary A. M. Rogers, and Patrick D. Schloss. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Medicine*, 8(1), apr 2016. doi: 10.1186/s13073-016-0290-3. URL https://doi.org/10.1186/s13073-016-0290-3.

[19] J. P. Zackular, M. A. M. Rogers, M. T. Ruffin, and P. D. Schloss. The human gut microbiome as a screening tool for colorectal cancer. *Cancer Prevention Research*, 7(11):1112–1121, aug 2014. doi: 10.1158/1940-6207.capr-14-0129. URL https://doi.org/10.1158/1940-6207.capr-14-0129.

[20] Weiguang Chen, Fanlong Liu, Zongxin Ling, Xiaojuan Tong, and Charlie Xiang. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS ONE*, 7(6):e39743, jun 2012. doi: 10.1371/journal.pone.0039743. URL https://doi.org/10.1371/journal.pone.0039743.

[21] Dirk Gevers, Subra Kugathasan, Lee A. Denson, Yoshiki Vazquez-Baeza, Will Van Treuren, Boyu Ren, Emma Schwager, Dan Knights, Se Jin Song, Moran Yassour, Xochitl C. Morgan, Aleksandar D. Kostic, Chengwei Luo, Antonio González, Daniel McDonald, Yael Haberman, Thomas Walters, Susan Baker, Joel Rosh, Michael Stephens, Melvin Heyman, James Markowitz, Robert Baldassano, Anne Griffiths, Francisco Sylvester, David Mack, Sandra Kim, Wallace Crandall, Jeffrey Hyams, Curtis Huttenhower, Rob Knight, and Ramnik J. Xavier. The treatment-naive microbiome in new-onset crohn's disease. *Cell Host & Microbe*, 15(3):382–392, mar 2014. doi: 10.1016/j.chom.2014.02.005. URL https://doi.org/10.1016/j.chom.2014.02.005.

[22] Eliseo Papa, Michael Docktor, Christopher Smillie, Sarah Weber, Sarah P. Preheim, Dirk Gevers, Georgia Giannoukos, Dawn Ciulla, Diana Tabbaa, Jay Ingram, David B. Schauer, Doyle V. Ward, Joshua R. Korzenik, Ramnik J. Xavier, Athos Bousvaros, and Eric J. Alm. Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease. *PLoS ONE*, 7(6):e39242, jun 2012. doi: 10.1371/journal.pone.0039242. URL https://doi.org/10.1371/journal.pone.0039242.

[23] Xochitl C Morgan, Timothy L Tickle, Harry Sokol, Dirk Gevers, Kathryn L Devaney, Doyle V Ward, Joshua A Reyes, Samir A Shah, Neal LeLeiko, Scott B Snapper, Athos Bousvaros, Joshua Korzenik, Bruce E Sands, Ramnik J Xavier, and Curtis Huttenhower. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biology*, 13(9):R79, 2012. doi: 10.1186/gb-2012-13-9-r79. URL https://doi.org/10.1186/gb-2012-13-9-r79.

[24] Ben P. Willing, Johan Dicksved, Jonas Halfvarson, Anders F. Andersson, Marianna Lucio, Zongli Zheng, Gunnar Järnerot, Curt Tysk, Janet K. Jansson, and Lars Engstrand. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology*, 139(6):1844–1854.e1, dec 2010. doi: 10.1053/j.gastro.2010.08.049. URL https://doi.org/10.1053/j.gastro.2010.08.049.

[25] Peter J. Turnbaugh, Micah Hamady, Tanya Yatsunenko, Brandi L. Cantarel, Alexis Duncan, Ruth E. Ley, Mitchell L. Sogin, William J. Jones, Bruce A. Roe, Jason P. Affourtit, Michael Egholm, Bernard Henrissat, Andrew C. Heath, Rob Knight, and Jeffrey I. Gordon. A core gut microbiome in obese and lean twins. *Nature*, 457(7228):480–484, nov 2008. doi: 10.1038/nature07540. URL https://doi.org/10.1038/nature07540.

[26] Juan S Escobar, Bernadette Klotz, Beatriz E Valdes, and Gloria M Agudelo. The gut microbiota of colombians differs from that of americans, europeans and asians. *BMC Microbiology*, 14(1), dec 2014. doi: 10.1186/s12866-014-0311-6. URL https://doi.org/10.1186/s12866-014-0311-6.

[27] Margaret L. Zupancic, Brandi L. Cantarel, Zhenqiu Liu, Elliott F. Drabek, Kathleen A. Ryan, Shana Cirimotich, Cheron Jones, Rob Knight, William A. Walters, Daniel Knights, Emmanuel F. Mongodin, Richard B. Horenstein, Braxton D. Mitchell, Nanette Steinle, Soren Snitker, Alan R. Shuldiner, and Claire M. Fraser. Analysis of the gut microbiota in the old order amish and its relation to the metabolic syndrome. *PLoS ONE*, 7(8):e43052, aug 2012. doi: 10.1371/journal.pone.0043052. URL https://doi.org/10.1371/journal.pone.0043052.

[28] Matthew C Ross, Donna M Muzny, Joseph B McCormick, Richard A Gibbs, Susan P Fisher-Hoch, and Joseph F Petrosino. 16s gut community of the cameron county hispanic cohort. *Microbiome*, 3(1):7, 2015. doi: 10.1186/s40168-015-0072-y. URL https://doi.org/10.1186/s40168-015-0072-y.

[29] Julia K. Goodrich, Jillian L. Waters, Angela C. Poole, Jessica L. Sutter, Omry Koren, Ran Blekhman, Michelle Beaumont, William Van Treuren, Rob Knight, Jordana T. Bell, Timothy D. Spector, Andrew G. Clark, and Ruth E. Ley. Human genetics shape the gut microbiome. *Cell*, 159(4):789–799, nov 2014. doi: 10.1016/j.cell.2014.09.053. URL https://doi.org/10.1016/j.cell.2014.09.053.

[30] David Zeevi, Tal Korem, Niv Zmora, David Israeli, Daphna Rothschild, Adina Weinberger, Orly Ben-Yacov, Dar Lador, Tali Avnit-Sagi, Maya Lotan-Pompan, Jotham Suez, Jemal Ali Mahdi, Elad Matot, Gal Malka, Noa Kosower, Michal Rein, Gili Zilberman-Schapira, Lenka Dohnalová, Meirav Pevsner-Fischer, Rony Bikovsky, Zamir Halpern, Eran Elinav, and Eran Segal. Personalized nutrition by prediction of glycemic responses. *Cell*, 163(5):1079–1094, nov 2015. doi: 10.1016/j.cell.2015.11.001. URL https://doi.org/10.1016/j.cell.2015.11.001.

[31] G. D. Wu, J. Chen, C. Hoffmann, K. Bittinger, Y.-Y. Chen, S. A. Keilbaugh, M. Bewtra, D. Knights, W. A. Walters, R. Knight, R. Sinha, E. Gilroy, K. Gupta, R. Baldassano, L. Nessel, H. Li, F. D. Bushman, and J. D. Lewis. Linking long-term dietary patterns with gut microbial enterotypes. *Science*, 334(6052):105–108, sep 2011. doi: 10.1126/science.1208344. URL https://doi.org/10.1126/science.1208344.

[32] A. M. Schubert, M. A. M. Rogers, C. Ring, J. Mogle, J. P. Petrosino, V. B. Young, D. M. Aronoff, and P. D. Schloss. Microbiome data distinguish patients with clostridium difficile infection and non-c. difficile-associated diarrhea from healthy controls. *mBio*, 5(3), may 2014. doi: 10.1128/mbio.01021-14. URL https://doi.org/10.1128/mbio.01021-14.

[33] Reiner Jumpertz, Duc Son Le, Peter J Turnbaugh, Cathy Trinidad, Clifton Bogardus, Jeffrey I Gordon, and Jonathan Krakoff. Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. *The American Journal of Clinical Nutrition*, 94(1):58–65, may 2011. doi: 10.3945/ajcn.110.010132. URL https://doi.org/10.3945/ajcn.110.010132.

[34] Marc A. Sze and Patrick D. Schloss. Looking for a signal in the noise: Revisiting obesity and the microbiome. *mBio*, 7(4), aug 2016. doi: 10.1128/mbio.01018-16. URL https://doi.org/10.1128/mbio.01018-16.

[35] Caroline Vincent, David A Stephens, Vivian G Loo, Thaddeus J Edens, Marcel A Behr, Ken Dewar, and Amee R Manges. Reductions in intestinal clostridiales precede the development of nosocomial clostridium difficile infection. *Microbiome*, 1(1):18, 2013. doi: 10.1186/2049-2618-1-18. URL https://doi.org/10.1186/2049-2618-1-18.

[36] G. Zeller, J. Tap, A. Y. Voigt, S. Sunagawa, J. R. Kultima, P. I. Costea, A. Amiot, J. Bohm, F. Brunetti, N. Habermann, R. Hercog, M. Koch, A. Luciani, D. R. Mende, M. A. Schneider, P. Schrotz-King, C. Tournigand, J. Tran Van Nhieu, T. Yamada, J. Zimmermann, V. Benes, M. Kloor, C. M. Ulrich, M. von Knebel Doeberitz, I. Sobhani, and P. Bork. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Molecular Systems Biology*, 10(11): 766–766, nov 2014. doi: 10.15252/msb.20145645. URL https://doi.org/10.15252/msb.20145645.

[37] Robert C. Edgar. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19):2460–2461, aug 2010. doi: 10.1093/bioinformatics/btq461. URL https://doi.org/10.1093/bioinformatics/btq461.

[38] Q. Wang, G. M. Garrity, J. M. Tiedje, and J. R. Cole. Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73(16):5261–5267, jun 2007. doi: 10.1128/aem.00062-07. URL https://doi.org/10.1128/aem.00062-07.

[39] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of machine learning research*, pages 2825–2830, 2011.

[40] Eric Jones, Travis Oliphant, and Pearu Peterson. SciPy: Open source scientific tools for python, 2001âĂŞ-, 2007.

[41] Statsmodels: Econometric and statistical modeling with python, 2010.

[42] J Oksanen. Multivariate analysis of ecological communities in r: vegan tutorial, 2011.

[43] Harold Tjalsma, Annemarie Boleij, Julian. R. Marchesi, and Bas E. Dutilh. A bacterial driver–passenger model for colorectal cancer: beyond the usual suspects. *Nature Reviews Microbiology*, 10(8):575–582, jun 2012. doi: 10.1038/nrmicro2819. URL https://doi.org/10.1038/nrmicro2819.

[44] Lixin Zhu, Robert D. Baker, and Susan S. Baker. Gut microbiome and nonalcoholic fatty liver diseases. *Pediatric Research*, 77(1-2):245–251, oct 2014. doi: 10.1038/pr.2014.157. URL https://doi.org/10.1038/pr.2014.157.

[45] Tingting Wang, Guoxiang Cai, Yunping Qiu, Na Fei, Menghui Zhang, Xiaoyan Pang, Wei Jia, Sanjun Cai, and Liping Zhao. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *The ISME Journal*, 6(2):320–329, aug 2011. doi: 10.1038/ismej.2011.109. URL https://doi.org/10.1038/ismej.2011.109.

[46] Gene V. Glass. Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10):3–8, nov 1976. doi: 10.3102/0013189x005010003. URL https://doi.org/10.3102/0013189x005010003.

[47] Francislon S Oliveira, John Brestelli, Shon Cade, Jie Zheng, John Iodice, Steve Fischer, Cristina Aurrecoechea, Jessica C Kissinger, Brian P Brunk, Christian J Stoeckert, Gabriel R Fernandes, David S Roos, and Daniel P Beiting. MicrobiomeDB: a systems biology platform for integrating, mining and analyzing microbiome experiments. *Nucleic Acids Research*, 46(D1):D684–D691, nov 2017. doi: 10.1093/nar/gkx1027. URL https://doi.org/10.1093/nar/gkx1027.

[48] Susan M. Huse, Les Dethlefsen, Julie A. Huber, David Mark Welch, David A. Relman, and Mitchell L. Sogin. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genetics*, 4(11):e1000255, nov 2008. doi: 10.1371/journal.pgen.1000255. URL https://doi.org/10.1371/journal.pgen.1000255.

[49] Omry Koren, Dan Knights, Antonio Gonzalez, Levi Waldron, Nicola Segata, Rob Knight, Curtis Huttenhower, and Ruth E. Ley. A guide to enterotypes across the human body: Meta-analysis of microbial community structures in human microbiome datasets. *PLoS Computational Biology*, 9(1):e1002863, jan 2013. doi: 10.1371/journal.pcbi.1002863. URL https://doi.org/10.1371/journal.pcbi.1002863.

[50] Lukasz Krych, Camilla H. F. Hansen, Axel K. Hansen, Frans W. J. van den Berg, and Dennis S. Nielsen. Quantitatively different, yet qualitatively alike: A meta-analysis of the mouse core gut microbiome with a view towards the human gut microbiome. *PLoS ONE*, 8(5):e62578, may 2013. doi: 10.1371/journal.pone. 0062578. URL https://doi.org/10.1371/journal.pone.0062578.

[51] Catherine A. Lozupone, Marcella Li, Thomas B. Campbell, Sonia C. Flores, Derek Linderman, Matthew J. Gebert, Rob Knight, Andrew P. Fontenot, and Brent E. Palmer. Alterations in the gut microbiota associated with HIV-1 infection. *Cell Host & Microbe*, 14(3):329–339, sep 2013. doi: 10.1016/j.chom.2013. 08.006. URL https://doi.org/10.1016/j.chom.2013.08.006.

[52] C. A. Lozupone, J. Stombaugh, A. Gonzalez, G. Ackermann, D. Wendel, Y. Vazquez-Baeza, J. K. Jansson, J. I. Gordon, and R. Knight. Meta-analyses of studies of the human microbiota. *Genome Research*, 23(10):1704–1714, jul 2013. doi: 10.1101/gr.151803.112. URL https://doi.org/10.1101/gr.151803.112.

[53] William A. Walters, Zech Xu, and Rob Knight. Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Letters*, 588(22):4223–4233, oct 2014. doi: 10.1016/j.febslet.2014.09.039. URL https://doi.org/10.1016/j.febslet.2014.09.039.

[54] Edoardo Pasolli, Duy Tin Truong, Faizan Malik, Levi Waldron, and Nicola Segata. Machine learning meta-analysis of large metagenomic datasets: Tools and biological insights. *PLOS Computational Biology*, 12(7):e1004977, jul 2016. doi: 10.1371/journal.pcbi.1004977. URL https://doi.org/10.1371/journal.pcbi.1004977.

[55] Emily Vogtmann, Xing Hua, Georg Zeller, Shinichi Sunagawa, Anita Y. Voigt, Rajna Hercog, James J. Goedert, Jianxin Shi, Peer Bork, and Rashmi Sinha. Colorectal cancer and the human gut microbiome: Reproducibility with whole-genome shotgun sequencing. *PLOS ONE*, 11(5):e0155362, may 2016. doi: 10. 1371/journal.pone.0155362. URL https://doi.org/10.1371/journal.pone.0155362.

Chapter 5

Increasing fecal microbiota transplant clinical trial successes by applying rational donor selection

Claire Duvallet, Caroline Zellmer, Pratik Panchal, Shrish Budree, Majdi Osman, and Eric J Alm.

The contents of this chapter constitute a manuscript in preparation for submission.

The figures appear within the main text of the chapter, and supplementary figures are at the end of the chapter.

Abstract

Early clinical successes are driving enthusiasm for fecal microbiota transplantation (FMT), the transfer of healthy gut bacteria through whole stool, as emerging research is linking the microbiome to many different diseases. However, preliminary trials have yielded mixed results and suggest that heterogeneity in donor stool may play a role in patient response. Thus, clinical trials may fail because an ineffective donor was chosen rather than because FMT is not appropriate for the indication. Here, we describe a conceptual framework to guide rational donor selection to increase the likelihood that FMT clinical trials will succeed. We argue that the mechanism by which the microbiome is hypothesized to be associated with a given indication should inform how donors are selected for FMT trials, categorizing these mechanisms into four disease models and presenting associated donor selection strategies. We next walk through examples based on previously published FMT trials and ongoing investigations to illustrate how donor selection might occur in practice. Finally, we show that typical FMT trials are not powered to discover individual taxa mediating patient responses, suggesting that clinicians should develop targeted hypotheses for retrospective analyses and design their clinical trials accordingly. Moving forward, developing and applying novel clinical trial design methodologies like rational donor selection will be necessary to ensure that FMT successfully translates into clinical impact.

5.1 Introduction

Fecal Microbiota Transplantation (FMT) is the transfer of gut bacteria through whole stool from a healthy donor to a recipient. FMT has demonstrated high cure rates in C. difficile infection (CDI) across multiple randomized, placebo-controlled trials [1] and has now entered standard of care for recurrent CDI in European and North American guidelines [2, 3, 4]. Beyond CDI, FMT is being explored in range of microbiome mediated diseases, and has demonstrated promising results in inflammatory bowel disease [5, 6, 7, 8, 9, 10].

Despite these early successes, the underlying mechanism of FMT across all disease indications, including CDI, remains unclear. However, it is generally considered that FMT restores gut microbial community perturbations from a dysbiotic to healthy stable state with engraftment of donor strains. Other donor-dependent features may be important such as the abundance of non-bacterial components or donor clinical features [11, 12]. However, not all FMT donors are alike: gut microbiota compositions vary significantly within healthy populations in ways that could impact the findings from an FMT trial [13]. This critical point of donor microbiome variation is rarely considered in the development of FMT trials [14, 15].

Unlike FMT trials in CDI, where selecting donors based on specific clinical or microbiome profiles does not seem to affect clinical response rates, donor selection is likely to be crucial to trial outcomes in diseases with more complex host-microbiome interplay or distinct disease-associated perturbations. Most notably, in a randomized controlled trial (RCT) of FMT for ulcerative colitis (UC) using 5 donors, 78% of patients who achieved remission after FMT received stool from a single donor [16]. Without this single donor the trial would have returned a negative result. Given the variation in donor microbiomes and donors' potential impact on clinical efficacy, how should clinicians and investigators select their donors for a clinical trial?

To date, the typical approach for donor selection in FMT trials is to use a single healthy screened donor or to randomly select multiple donors from a set of screened potential donors [10, 17, 18]. However, in clinical indications where successful donors

may be rare, such as UC, clinical trials with randomly-selected donors may fail not because FMT is inappropriate for the indication, but because an ineffective donor was chosen. An alternative approach is to expose patients to multiple donors in order to mitigate the risk of sub-optimal donor selection. In a large RCT of FMT in UC, FMT enemas for a single patient were derived from between three and seven donors with patients receiving multiple donors throughout the 8 week course of treatment [10]. However, using multiple donors for a single patient may not be feasible or appropriate in many disease indications or clinical trial settings (e.g. single-dose FMT studies). Continuing the strategy of sub-optimal, random donor selection when it is not warranted risks returning false negative trials, stalling the field and delaying the development of novel therapies for seemingly intractable microbiome-mediated conditions.

Unlike traditional clinical trials which test well-defined small molecules, the therapy under study in FMT trials, the donor microbiome, also varies [15]. Thus, translational FMT research requires a paradigm shift in order to systematically address rational donor selection. Fortunately, with the emergence of large multi-donor stool banks, expanded access to genome sequencing technologies and publicly available microbiome sequencing datasets, rational donor selection is feasible and presents a unique opportunity to advance the research methods of this nascent field.

In this paper, we present a framework to guide donor selection for FMT trials. The mechanism by which the microbiome is hypothesized to be associated with a given indication should inform how donors are selected for FMT trials, and we describe different disease models which may underlie microbiome-mediated conditions (Figure 5-1). We describe strategies to rationally select donors for each type of disease model, and provide examples based on previously published FMT trials and ongoing investigations. Finally, we discuss limitations of performing discovery-based retrospective research after an FMT clinical trial concludes. To our knowledge, this is the first description of a comprehensive framework for rational donor selection in FMT trials.

5.2 Models of microbiome-mediated disease

FMT trials are pursued when research or clinical experiences suggest that a condition may be causally linked to the microbiome. Here, we propose four different models which may underlie microbiome-mediated etiologies and their corresponding rational donor selection strategies (Figure 5-1). Ultimately, it is up to each individual clinician-researcher to determine which of these model(s) are relevant in their specific case, based on published cross-sectional studies, mechanistic investigations in model organisms, and their own clinical experience treating patients. Additionally, logistical considerations will be important factors in making the final donor selection regardless of which strategy is pursued. Clinicians should ensure that the pool of donors that they are screening have enough material to sustain the required number of FMTs for their entire trial.

Most of the donor selection strategies described below can be modified to incorporate matching between patients and donors. More specifically, donors can be tailored to individual patients to specifically make up for the unique taxonomic or functional deficiencies in that patient's microbiome. With the increasing amount of microbiome data available from published FMT trials, we encourage collaborations between clinicians and bioinformaticians to analyze these data in order to generate or perhaps even confirm the validity of potential donor selection strategies before selecting one (Figure 5-2). Finally, the strategies presented here should also be combined with adaptive clinical trial designs to further increase the probability of having a successful FMT trial [19].

5.2.1 Acute dysbiosis

An acutely dysbiotic gut microbial community is broadly dysfunctional and can no longer maintain the health of the host. For example, in the case of recurrent Clostridium difficile infection, a disturbed microbial community is unable to prevent colonization by the pathogen, leading to recurrent overgrowth of C. diff and clinical symptoms [20]. Acute dysbiosis has also been described with the "Anna Karenina principle": all
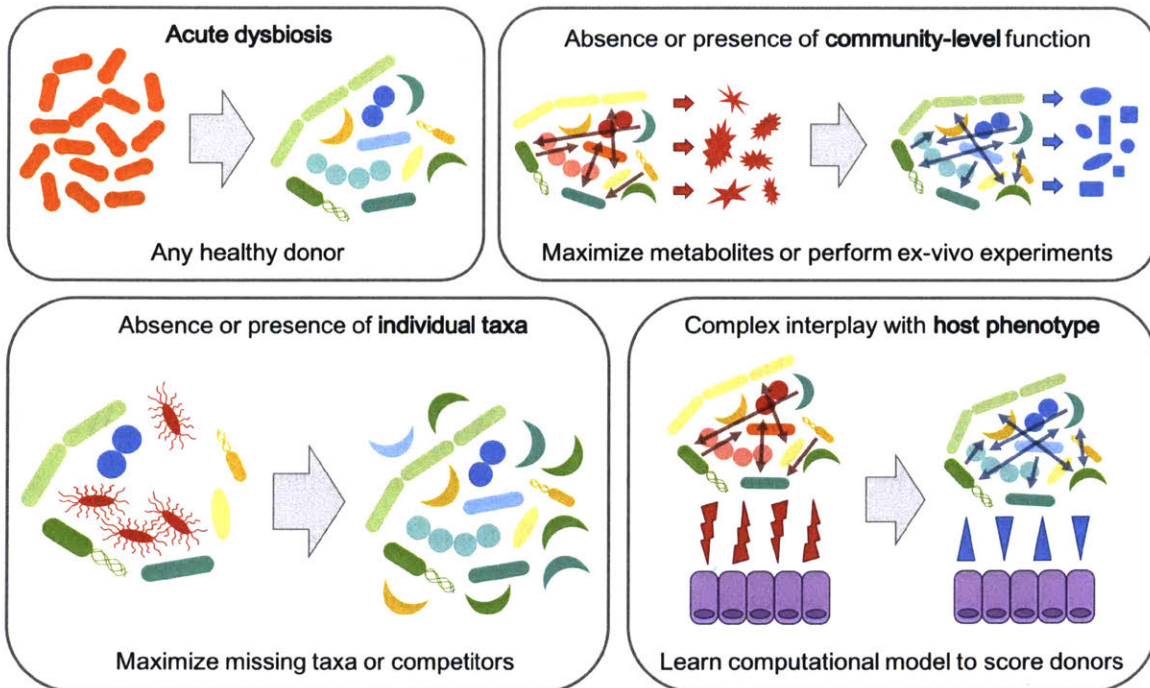
116

Figure 5-1: Overview of the different models of microbiome-mediated disease and associated donor selection strategy.

healthy microbiomes are alike but dysbiotic communities are all dysbiotic in their own ways [21]. In this view of acute dysbiosis, microbial communities respond stochastically to stressors, resulting in dysbiotic communities which are characterized by increased variability rather than deterministic shifts to precise community type(s) [21].

In this model, the host just needs to return to a "healthy" microbiome and thus choosing any healthy FMT donor should be sufficient to induce clinical improvements. Because there is no specific disease-associated microbial community and deviation from health is instead the more important factor, simply replenishing the microbiome with a healthy configuration should be sufficient. Indeed, FMT trials have demonstrated that recurrent C. diff infection can be effectively treated by almost any choice of donor [22]. In this case, researchers should consider how they define a "healthy" microbiome and how they will ensure engraftment of the transplanted healthy communities.

5.2.2 Absence or presence of individual taxa

Absence of beneficial taxa

In other cases, perhaps a disease is being caused or exacerbated by the lack of certain specific microbes, and replenishing these few taxa would be sufficient to restore the host to health. For example, Hsiao et al showed that a single microbe, *R. obeum*, restricted infection by *V. cholerae* through quorum-sensing-mediated mechanisms [23]. Surprisingly, non-communicable diseases may also fall into this model: Wilck et al. demonstrated that a single strain of Lactobacillus was sufficient to prevent salt-induced hypertension, and follow-up studies indicate that similar mechanisms may be involved in salt-sensitive high blood pressure in humans as well [24].

In these cases, the donor selection strategy should focus on maximizing the probability of engraftment of the beneficial taxa. In cases where the unique taxa are not specifically known or are rare members of the human microbiota, many healthy donors should be pooled together to maximize the probability that the transplanted sample contains the necessary taxa. If the missing microbes are known and well-characterized, on the other hand, researchers can screen their pool of potential donors to find the sample with the highest abundance of these taxa.

Presence of harmful taxa

Rather than being characterized by the absence of individual bacteria, perhaps a disease is instead mediated by the presence or overabundance of specific microbes, and removing these bacteria in a targeted fashion could lead to improvements in disease progression. For example, *Fusobacterium* has been found to be more abundant in colorectal cancer patients, specifically enriched in the tumors themselves [25]. Multiple groups have identified mechanistic associations between *Fusobacterium*, inflammatory transcriptional signatures, and tumor growth in mouse and human models of colorectal cancer, pointing to a causal role for *Fusobacterium* in colorectal cancer progression [25, 26]. Recent work has found that treating tumors with antibiotics slows tumor progression, further confirming these causal associations and pointing

toward potential microbiome-based therapeutic interventions [27].

Removing and replacing these bacteria should be the goal of FMT in cases where this disease model applies. This can be achieved by first removing the harmful bacteria in a targeted way (e.g. via antibiotic treatment) with follow-up FMT to re-establish a healthy community that prevents their re-colonization. In all cases, donors should be screened to exclude any samples which contain the harmful bacteria. Donor samples can then be selected based on the abundance of bacteria which are known to out-compete the harmful taxa. Competitors can be identified by searching the microbiology literature to identify bacteria which live in the same niche or which have been experimentally shown to directly out-compete the undesirable taxa, or they can perform these competition assays themselves. If resources to perform competition assays are not available and the literature is sparse, researchers can also mine existing microbiome data to find bacteria which consistently anti-correlate with the harmful taxa, and choose donor samples with a high abundance of these putative competitors.

Patient matching

Taxa-based donor selection strategies are particularly amenable to patient-matching, when both patient and donor microbiome data are available prior to the start of a trial. For example, if one patient is completely missing some of the beneficial taxa but not others, then these taxa can be weighted more heavily in the donor selection process. The phylogenetic relationships between donor and recipient taxa could also be incorporated into donor selection: if a patient already has many bacteria which are closely phylogenetically related to known competitors of some of the harmful bacteria, then competitors of the other harmful bacteria can be upweighted in the donor selection process. Similarly, if patients already have taxa which are already filling certain niches important for health, the taxa which fill those same niches can be downweighted in donor selection.

119

Case study: Inflammatory Bowel Disease

An example where the "missing taxa" model may be applicable is in inflammatory bowel disease (IBD). Butyrate has long been associated with inflammatory bowel disease [28], and recent case-control and longitudinal studies point to a consistent lack of butyrate-producing bacteria in patients with IBD [29, 30]. Furthermore, preliminary FMT trials in IBD have been marked by significant donor variability and suggest that donor microbiome characteristics may be associated with FMT response [16, 31]. These results indicate that IBD may benefit from rational donor selection approach, and that donors with high abundances of butyrate-producing organisms may yield higher FMT response rates than randomly selected donors.

Given the availability of microbiome data from completed FMT studies, we tested this hypothesis that IBD trials would benefit from a rational donor selection strategy based on the "absence of beneficial taxa" disease model. We re-analyzed microbiome data from three completed IBD FMT trials which provided publicly available sequencing data for patient and donor samples [31, 32, 33]. We selected butyrate-producers based on their genus-level taxonomy, using the genera identified in Vital et al. 2017 (see Methods, [34]). Donors in the three studies exhibited a range of total abundances of butyrate-producing bacteria (Figure 5-2A). Surprisingly, however, the abundance of butyrate producers in the donor stool was not associated with recipient patients' clinical responses (Figure 5-2B). We also found no association with response when matching donor abundances with their respective patient's original abundance of butyrate producers (Supplementary Figure 5-5). These results show that selecting donors based on the abundance of butyrate producers may not yield improved clinical trial outcomes in IBD, and illustrates the process by which clinicians could approach and validate a rational donor selection strategy based on individual taxa. More complex methods to identify butyrate producers (e.g. using phylogenetic-aware methods and/or metagenomics data) could be used in the next iteration to develop a donor selection strategy, if these data are available to clinicians. Another approach, discussed below, is to select donors based on functional community assays and direct
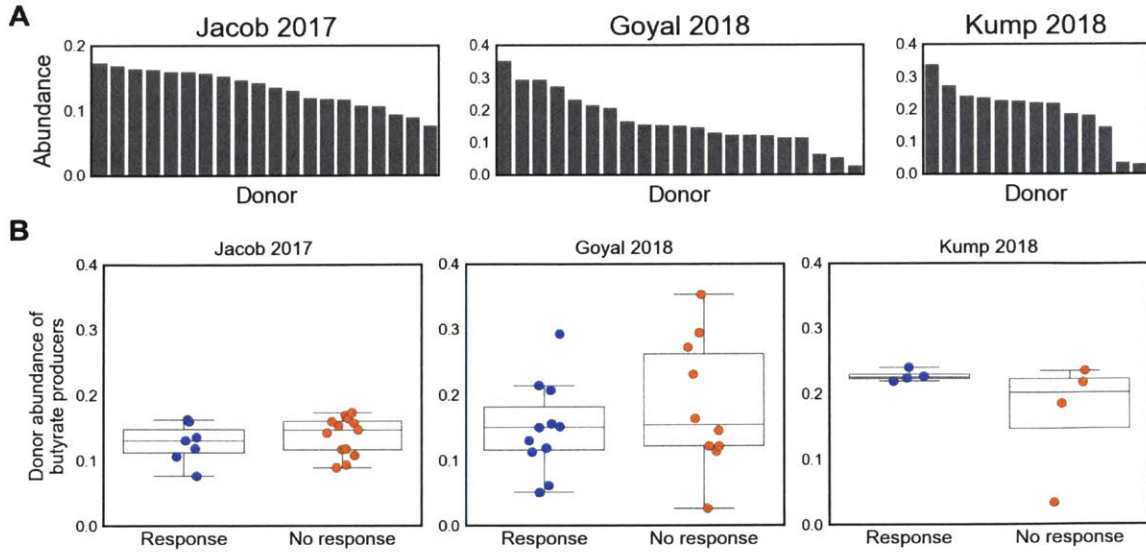
120

Figure 5-2: Case study in IBD: select donors based on abundance of butyrate producers? (A) abundance of butyrate producers in each study's donor samples. (B) abundance of butyrate producers in donor samples, stratified by respective patient's response.

measurement of butyrate production rather than microbial taxonomies alone.

5.2.3 Community-level functionality

Some microbiome-associated diseases may not be addressable by replenishing the patient with a generically healthy community or by targeting individual taxa, and may instead be mediated by the microbiome through a community-level function. Here, there may not be a consistent disease-associated microbiome across patients in terms of taxonomic composition, but patients may be characterized by having microbiomes which are similarly missing or enriched in some core functionality. This model may also apply to conditions where there are consistent disease- or health-associated taxa, but in which their collective functioning is the more important mediator of disease. The IBD case study described above may reflect this situation: although depletion of butyrate producers is strongly associated with IBD throughout the literature, a successful donor selection strategy may need to consider butyrate production directly rather than through the proxy of taxonomy [29, 30].

121

Missing community-level function

In the case where a community-level function is missing from patients' microbiomes, the goal of FMT should be to replace the deficient community with a beneficially functional microbiome. Here, it is important that a single donor with an intact microbial community is used, rather than a mixture of donors which may not yield the desirable community composition at steady-state after engraftment. To choose a donor, molecules which can serve as proxies for the metabolic output of the microbial community can be measured directly in donor stool, and donors can be selected based on the abundance of these molecules.

Like IBD, hepatic encephalopathy (HE) is an example where community functionality is likely more relevant to FMT outcome than specific taxa. A previous trial in HE [35] rationally selected their single donor by maximizing the abundance of *Lachnospiraceae* and *Ruminococcaceae*, taxa which were identified based on cross-sectional microbiome data. The clinical trial was a success, but it remains unclear from this trial whether the donor's strains engrafted in the patients post-FMT and whether this played any role in the successful FMT responses. The exact mechanisms of action of these strains remains unknown, though both bacterial families are known short chain fatty acid producers (in particular butyrate) [34]. More recent studies have more directly implicated the production of short chain fatty acids and secondary bile acids as being important in liver cirrhosis and subsequent complications such as HE, suggesting that community-level functioning may be a more important driver of FMT response. Thus, HE may be a case in which function-based donor selection can be employed.

To illustrate this process, we analyzed stool metabolomics data from 83 Open-Biome donors and used this data to rank them based on their estimated production of short chain fatty acids and secondary bile acids (Figure 5-3). As in the IBD case study, we found that donors exhibited a range of values for our metabolites of interest (Figure 5-3A and C). We ranked donors based on their amounts of the three measured short-chain fatty acids (butyrate, isovalerate, and propionate) and on their

bile acid conversion rates, approximated as the ratio between the total amounts of primary and secondary bile acids (Figure 5-3B and D). With this process, we were able to identify four donors who were in the top 25% of all donors for both metrics (Figure 5-3E). In a real FMT trial, a clinician would then work with their stool bank to ensure that these donors were still active and/or had enough material to be used in the full trial.

While measuring metabolites in stool as a proxy for community production will likely be an improvement over taxonomy-based approaches in most cases, these measurements are also complicated by potential host effects. For example, different hosts may absorb these molecules at different rates, and so measuring them in stool may not be an accurate reflection of each donor community's productive potential. Additionally, community function may depend on non-biologically relevant factors like the donor's diet and time that they provided their sample. As an example, bile acid production spikes after meals [36], so the amount of bile acids measured in a given stool sample may reflect the amount of time since the donor last ate rather than their actual microbial community's functional production of these molecules. If clinicians have access to sufficient resources, a better way to screen donors may be to perform ex-vivo assays, in which each donor sample is homogenized and provided with the substrates (e.g. fiber) needed to produce the desirable output (e.g. short-chain fatty acids like butyrate). In this way, the donor community function can be measured directly [37, 38].

Overactive function

A disease may also be mediated by an overactive microbiome doing something harmful to the host. For example, many studies have shown a causal association between TMAO produced by the microbiota and atherosclerosis [39, 40]. Here, the goal of FMT should also be to replace the patient's microbiome with a beneficially functional community, but the donor selection strategy may attempt to identify communities in which the harmful function is completely absent or which produces an inhibitor of the harmful microbe-derived molecule [40].
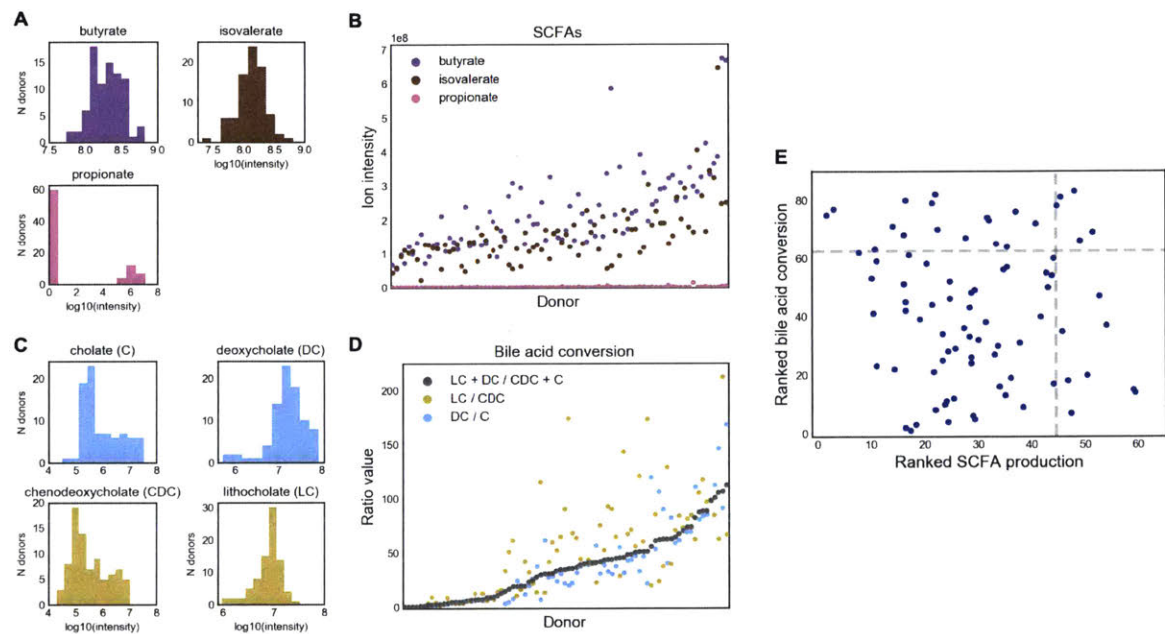
Figure 5-3: Case study in liver cirrhosis: selecting donors based on community function by mining stool metabolomics data. (A) Distribution of SCFAs in all donor stools. (B) Abundance of each SCFA per donor, ranked by average SCFA abundance. (C) Distribution of bile acids in all donors. Primary bile acids are in the left column, secondary bile acids are in the right column. Bile acids are colored according to pathways. (D) Bile acid conversion ratios in each donor, ranked by total secondary to primary bile acids. (E) The five donors in the top 25% for both of these metrics, for example, could be used in a rationally-designed liver cirrhosis FMT trial.

5.2.4 Microbiome-associated host phenotypes

Diseases with more complex etiologies may not have a direct taxonomic or functional association with the microbiome but instead be related through some intermediate host phenotype which needs to be improved or corrected. For example, severe acute malnutrition has been associated with a gut microbiota which is not fully mature, with mouse experiments suggesting that this association may be causal [41, 42]. Other studies have shown a relationship between gut microbiome, immune development, and development of autoimmune conditions later in life [43, 44, 45]. These relationships may have mechanistic explanations which are not directly measurable from donor or patient stool (e.g. immunogenicity of bacteria, ability of bacteria to eat the host's mucus, etc) but which can nevertheless be inferred from existing data and used to select potential donors.

For these more complex cases, models can be trained from existing datasets to learn the community signatures linked to the disease-associated phenotype. In some cases, it may be possible to develop computational models which directly predict the phenotype of interest. For example, Stein et al. developed a model to predict the induction of regulatory T-cells by microbial communities [46]. In other cases with few known mechanistic models, machine learning algorithms can be trained on multiple cross-sectional datasets to identify complex signatures that reproducibly distinguish patients from healthy controls. These models can then be applied to score potential donors, and the donor with the "most healthy" score may be chosen for a trial.

5.3 Little understanding of underlying disease model

In some conditions, there may not be enough understanding of the underlying microbiome-based etiology to inform donor selection in an FMT trial. It may also be the case that there are no existing datasets on which to train models, existing datasets are not sufficiently powered to distinguish the different potential underlying models, or logistical considerations constrain a clinician's ability to select specific donors for their trial. In these cases, we recommend selecting different healthy donors, employing an adaptive

clinical trial design in which donors are cycled after they have clinical failures (as described previously in [19]), and performing retrospective analyses to answer targeted hypotheses which were developed during the clinical trial design process.

5.3.1 Cycling healthy donors in adaptive trials

As donors change through the course of an adaptive trial, clinicians may elect to select their donors randomly or to more rationally cycle through donors [19]. "Differently healthy" donors may be selected, perhaps representing different underlying disease-associated models described above. Donors may also be selected to span the range of "healthy" microbiomes in a given population. For example, clinicians may pick a "median" health donor (i.e. one who is about as similar to all reference microbiomes), define a "healthy plane" and pick donors based on their distance to this plane (as in [47]), or simply based on the presence or abundance of certain consistently "core" health-associated bacteria [48],[29]. In a similar vein, "healthy" donors can also be chosen based on their distance from disease-associated microbiomes, as opposed or in addition to similarity to health. Published case-control datasets can be used to identify donors with communities which are farthest away from the median or average diseased patient. These datasets can also be mined to identify taxa which are consistently disease-associated, and which should be minimized or perhaps even absent in the potential donor. Pairing rational donor selection with adaptive trial designs may eventually yield insight into the underlying model mediating the disease of interest if certain types of "healthy" donors consistently perform better at treating patients than others.

5.3.2 Discovery-based retrospective analyses

In these exploratory FMT clinical trials, discovering microbiome characteristics which are differentially associated with FMT response may be a valuable secondary endpoint [15], identifying characteristics of good donors and informing donor selection strategy for future trials. Furthermore, companies attempting to develop microbiome-based

126

therapeutics may use FMT trials to discover the key bacteria which mediate FMT response in order to include these in their microbial cocktail product. However, exploratory FMT trials tend to enroll few patients, limiting the potential power of retrospective analyses to find associations between the microbiome and FMT response.

We performed a simulation to determine the likelihood of a retrospective analysis to identify donor-derived bacteria associated with different patient responses to FMT. We performed this simulation for multiple FMT trial set-ups and outcomes (i.e. number of FMT responders and non-responders). We used existing microbiome datasets to model different effect sizes, where we use "effect size" to mean the number of bacteria which are differentially abundant in donor samples given to patients who did and did not respond to FMT. We used case-control datasets to model the microbiome data and various effect sizes, with a large effect represented by an infectious diarrhea dataset [49], a medium effect represented by colorectal cancer [50], and a weak effect represented by obesity [51]. For each of these datasets, we identified the top ten most differentially abundant bacteria in the overall population as the key mediating bacteria (see Methods). Next, we simulated different trials, varying the numbers of patients in the FMT arm and the FMT response rates (i.e. proportion of patients which were FMT responders, represented by sampling from the "case" patients, vs. non-responders, represented by sampling from the "control" patients, representing non-responders). We subsampled patients according to these parameter settings, identified differentially abundant genera, and compared these to the top ten genera identified from the entire datasets (Figure 5-4).

In cases where the microbial signature for FMT response is expected to be large (i.e. the difference in donor stools given to FMT responders vs. non-responders is as large as the effect of diarrhea effect on the microbiome), we found that small FMT trials would recover most of the top hits in most cases. The power to detect associations decreased as FMT response rates became less balanced (i.e. response rates different from 50%), and in these cases trials would need to include up to 50 patients in the FMT arm to recover the key mediating taxa. For both medium and small effect sizes, however, prohibitively large FMT arms would be needed to recover
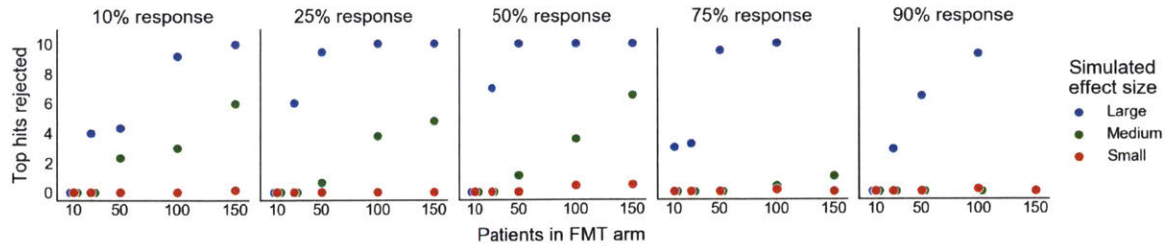
Figure 5-4: Power simulation results, showing how many of the 10 most "truly" differentially abundant genera would be recovered as significant under different FMT study designs. Each panel represents a different FMT response rate (i.e. percent of patients in the responder vs. non-responder group). The effect size (i.e. number of genera which are differentially abundant in FMT responders vs. non-responders) was simulated by using three different case-control microbiome datasets. A large effect size is modeled by the effect of diarrhea on the microbiome, medium by colorectal cancer, and small by obesity. The top 10 "true" differentially abundant genera were identified by calculating their signal-to-noise ratios in the full original dataset (i.e. mean difference divided by the standard deviation).

most key mediating taxa. We found that when the microbial signature for FMT is equivalent to the effect of diseases like colorectal cancer on the microbiome, at least 100 patients are needed in the FMT arm to recover at least half of the most truly differentially abundant genera for most FMT trials.

These results suggest that successful secondary analyses of microbiome data from FMT trials will require either very large FMT arms, investigating more targeted hypotheses, or additional sample collections. For example, clinicians may consider pairing donor and patient samples or collecting longitudinal patient samples to increase power to make discoveries. They may also consider testing specific hypotheses developed before the trial, such as comparing the total abundance of butyrate producers between FMT responders and non-responders, or performing functional assays to measure specific metabolites thought to be associated with FMT response. On the other hand, researchers wishing to identify the key taxa to include in an FMT drug may consider pursuing clinical trials in which identifying these taxa is the primary endpoint, and power them accordingly.

5.4 Discussion

The framework presented here encourages clinicians to leverage their clinical experience, existing microbiome research and published datasets, and the increasing availability of screened donor stools to more efficiently translate microbiome-based interventions into clinical impact. Clinicians can apply their existing knowledge and a priori hypotheses to determine which microbiome-mediated disease model may underlie their indication of interest, and then select donors accordingly. By rationally choosing donors during the FMT trial design, clinicians will increase the likelihood of successful FMT trials in diseases in which donor heterogeneity affects patient response. Our power simulation analysis also suggests that specific plans for retrospective analyses of the microbiome data generated should be developed during trial design, with targeted hypotheses of interest and sample collection plans tailored accordingly. Otherwise, exploratory analyses are unlikely to make new discoveries from most FMT trials. Paired with adaptive clinical trial designs, FMT trials with rationally-selected donors will become an important tool in advancing translational microbiome research and clinical treatment to improve and save patient lives.

As FMT-specific clinical trial design methodologies become more developed, many additional questions will need to be addressed. Some of these key questions relate to choosing healthy donors: what defines a "healthy" donor, and when and how should that definition change? Some screening criteria, like excluding certain known pathogens, will likely apply to all donors. Beyond these, however, there is little consensus on what defines a "healthy" microbiome. Consider a clinician carrying out an FMT trial in an African setting: given that healthy Africans from across the continent are known to have more *Prevotella* than North Americans, it might be advisable to source donors locally to better match the expected healthy state of the patients [52, 13, 53]. But what if the local population has higher asymptomatic colonization rates of undesirable bacteria like opportunistic pathogens, resulting in usual screening criteria excluding many or all potential donors: should the criteria be adapted to allow for local donors, or should donors be sourced from a foreign stool

bank whose donors may not match the local population? To answer this question, more research will need to be carried out to understand differences between healthy microbiomes globally and their clinical implications [54, 55].

On the patient side, comorbidities, lifestyle, and dynamic disease manifestations present additional challenges and opportunities to improve donor selection and FMT clinical trial designs. How should comorbidities be incorporated into donor selection? Patients with multiple disease processes may be dominated by one disease model or may exhibit a combination of models, perhaps affecting which donors would be optimal for their specific case. For example, a person whose condition involves both acute dysbiosis and community-level dysfunction might respond well to any healthy donor, or may require a more complex combination of total community replacement along with enrichment for community function. Relatedly, diseases that change manifestations over time may benefit from employing different donor selection strategies over the course of a longitudinal FMT trial. Additionally, although there have been no serious adverse events related to FMT material in either clinical practice for rCDI or in clinical trials across adults or pediatrics, could some donors further reduce the probability of adverse events in at-risk patients? Finally, how should other sources of heterogeneity like lifestyle, diet, and medication usage be incorporated into rational donor selection? In cases where FMT is combined with other microbiome-targeted interventions like prebiotics or dietary changes, could some donors have synergistic effects with these paired interventions and lead to greater clinical success?

To ensure that FMT reaches its full potential to improve and save patient lives, clinicians should think critically about how their FMT trials can be designed for maximal impact. By applying new approaches like rational donor selection and adaptive trial designs, the number of trials which fail even though they could have succeeded will decrease. Furthermore, by developing targeted hypotheses, post-trial analysis plans, and associated sample collection schema alongside the core FMT trial design itself, the number of basic scientific discoveries that are made from each trial will significantly increase. As FMT expands beyond rCDI and microbiome-based therapeutics are developed to target a range of diseases, novel methods and approaches

tailored to the unique challenges and opportunities presented by FMT will be critical to ensuring the advancement of translational microbiome science and beneficial impact on patient lives.

5.5 Methods

5.5.1 Microbiome data processing

Raw fastq data files were downloaded from the European Nucleotide Archive using the following accession numbers: Jacob et al 2017, PRJNA388210; Goyal et al. 2018, PRJNA380944; and Kump et al. 2018, PRJEB11841. All data was processed using QIIME 2 (v. 2018.6.0, [56]). Briefly, data was imported into QIIME 2 as paired-end (Kump et al. 2018; Jacob et al. 2017) or single-end (Goyal et al. 2018) data, filtered based on sequence quality with `quality-filter q-score`, and denoised with deblur using `deblur denoise-16S` [57]. Representative sequences were assigned taxonomy using `feature-classifier classify-sklearn` with the GreenGenes-trained Naive Bayes classifier provided by QIIME 2 (gg-13-8-99-nb-classifier.qza) [58]. All data was exported to tab-delimited format and analyzed in Python 2.7.6.

5.5.2 Quantifying abundance of butyrate producers

We identified butyrate producers at the genus-level based on the analysis performed in Vital et al. 2017 [34]. These taxa were detected in >70% of individuals in Vital et al. 2017, are known butyrate producers (with a majority of the analyzed representative genomes containing known butyrate production pathways), and accounted for the majority of the total butyrate pathway abundances in human metagenomics data. We removed *E. ventriosum*, *E. hallii*, and *E. rectale* from our analyses as these species-level taxa do not comprise one genus with conserved butyrate production.

5.5.3 Stool metabolomics

Metabolomics data was generated as described in Poyet, Groussin, Gibbons et al. (in preparation) and downloaded after personal communication with the authors. For donors with multiple samples, we considered the mean metabolite abundances across all sampled time points. We identified three short chain fatty acids in the data (propionate, butyrate, and isovalerate) and the major primary (cholate and chenodeoxycholate) and secondary (deoxycholate and lithocholate) bile acids. Lithocholate was available for both C-18 negative and HILIC negative modes; we considered only the C-18 negative data to match the other bile acids. Bile acid conversion rates were calculated as in Kakiyama et al. 2013 [59]. Donors were ranked based on their average SCFA abundances and based on the total bile acid conversion ratio ((lithocholate + deoxycholate) / (chenodeoxycholate + cholate)).

5.5.4 Power simulation

We performed a simulation to determine the power of FMT trials to retrospectively find associations between donor bacterial abundances and FMT response. We used case-control gut microbiome datasets from MicrobiomeHD [29] to model different effect sizes for FMT response. Here, we use "effect size" to mean the number of genera which are differentially abundant between patients who respond to FMT vs. patients who do not. Per the results in MicrobiomeHD, we used infectious diarrhea to model a large effect [49], colorectal cancer to model a medium effect [50], and obesity to model a small effect [51]. We collapsed OTUs to genus-level as in [29] and ranked genera according to their signal-to-noise ratio in each entire dataset, where the signal-to-noise was calculated as the difference in mean log abundance in cases and controls divided by the standard deviation of the log abundances across all samples. We considered the 10 genera with the largest absolute signal-to-noise ratios as our "top hits" in the main text.

We modeled different FMT clinical trial designs and outcomes by varying the number of total patients in the trial and the percent of FMT responders (i.e. the

number of patients we selected from the original "case" group relative to the original "control" patients, to model FMT responders and non-responders). For each of these designs, we subsampled the correct number of case samples to represent FMT responders and control samples to represent non-responders from the original datasets. We identified significantly differentially abundant genera with the `kruskalwallis` function from `scipy.stats.mstats` (scipy v. 1.1.0) as genera with q < 0.05 after multiple hypothesis testing correction with the `multipletests` function (`method='fdr_bh'`) from the `statsmodels.sandbox.stats.multicomp` package (statsmodels v. 0.9.0). We then counted how many of the top genera identified through the signal-to-noise ranking were identified as significantly different as a proxy for the power to detect effects.

5.5.5   Code and data availability

Code to reproduce all of these analyses and figures can be found at `https://github.com/cduvallet/donor-selection/`. Data were downloaded from original sources as described above.
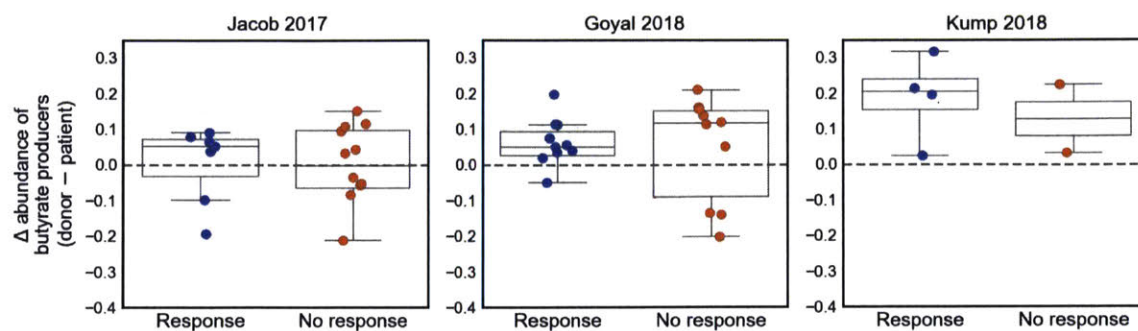
5.6 Supplementary Figure

Figure 5-5: Difference between abundance of butyrate producers in donor sample and respective patient sample, stratified by patient response.

Bibliography

[1] M. N. Quraishi, M. Widlak, N. Bhala, D. Moore, M. Price, N. Sharma, and T. H. Iqbal. Systematic review with meta-analysis: the efficacy of faecal microbiota transplantation for the treatment of recurrent and refractory clostridium difficile infection. *Alimentary Pharmacology & Therapeutics*, 46(5):479–493, jul 2017. doi: 10.1111/apt.14201. URL https://doi.org/10.1111/apt.14201.

[2] L Clifford McDonald, Dale N Gerding, Stuart Johnson, Johan S Bakken, Karen C Carroll, Susan E Coffin, Erik R Dubberke, Kevin W Garey, Carolyn V Gould, Ciaran Kelly, Vivian Loo, Julia Shaklee Sammons, Thomas J Sandora, and Mark H Wilcox. Clinical practice guidelines for clostridium difficile infection in adults and children: 2017 update by the infectious diseases society of america (IDSA) and society for healthcare epidemiology of america (SHEA). *Clinical Infectious Diseases*, 66(7):e1–e48, feb 2018. doi: 10.1093/cid/cix1085. URL https://doi.org/10.1093/cid/cix1085.

[3] Giovanni Cammarota, Gianluca Ianiro, Herbert Tilg, Mirjana Rajilić-Stojanović, Patrizia Kump, Reetta Satokari, Harry Sokol, Perttu Arkkila, Cristina Pintus, Ailsa Hart, Jonathan Segal, Marina Aloi, Luca Masucci, Antonio Molinaro, Franco Scaldaferri, Giovanni Gasbarrini, Antonio Lopez-Sanroman, Alexander Link, Pieter de Groot, Willem M de Vos, Christoph Högenauer, Peter Malfertheiner, Eero Mattila, Tomica Milosavljević, Max Nieuwdorp, Maurizio Sanguinetti, Magnus Simren, and Antonio Gasbarrini. European consensus conference on faecal microbiota transplantation in clinical practice. *Gut*, 66(4):569–580, jan 2017. doi: 10.1136/gutjnl-2016-313017. URL https://doi.org/10.1136/gutjnl-2016-313017.

[4] Christina M Surawicz, Lawrence J Brandt, David G Binion, Ashwin N Ananthakrishnan, Scott R Curry, Peter H Gilligan, Lynne V McFarland, Mark Mellow, and Brian S Zuckerbraun. Guidelines for diagnosis, treatment and prevention of clostridium difficile infections. *The American Journal of Gastroenterology*, 108(4):478–498, feb 2013. doi: 10.1038/ajg.2013.4. URL https://doi.org/10.1038/ajg.2013.4.

[5] Pratik Panchal, Shrish Budree, Alex Scheeler, Geraldine Medina, Monica Seng, Wing Fei Wong, Ryan Eliott, Thomas Mitchell, Zain Kassam, Jessica R. Allegretti, and Majdi Osman. Scaling safe access to fecal microbiota transplantation: Past, present, and future. *Current Gastroenterology Reports*, 20(4),

mar 2018. doi: 10.1007/s11894-018-0619-8. URL https://doi.org/10.1007/s11894-018-0619-8.

[6] Jeffrey Gelfand. Fecal microbiota transplantation (FMT) of FMP30 in relapsing-remitting multiple sclerosis. ClinicalTrials.gov, 2018. URL https://clinicaltrials.gov/ct2/show/NCT03594487. Accessed November 28, 2018.

[7] Ruud S. Kootte, Evgeni Levin, Jarkko Salojärvi, Loek P. Smits, Annick V. Hartstra, Shanti D. Udayappan, Gerben Hermes, Kristien E. Bouter, Annefleur M. Koopen, Jens J. Holst, Filip K. Knop, Ellen E. Blaak, Jing Zhao, Hauke Smidt, Amy C. Harms, Thomas Hankemeijer, Jacques J.G.H.M. Bergman, Hans A. Romijn, Frank G. Schaap, Steven W.M. Olde Damink, Mariette T. Ackermans, Geesje M. Dallinga-Thie, Erwin Zoetendal, Willem M. de Vos, Mireille J. Serlie, Erik S.G. Stroes, Albert K. Groen, and Max Nieuwdorp. Improvement of insulin sensitivity after lean donor feces in metabolic syndrome is driven by baseline intestinal microbiota composition. *Cell Metabolism*, 26(4):611–619.e6, oct 2017. doi: 10.1016/j.cmet.2017.09.008. URL https://doi.org/10.1016/j.cmet.2017.09.008.

[8] Majdi Osman. Transfer of healthy gut flora for restoration of intestinal microbiota via enema for severe acute malnutrition. ClinicalTrials.gov, 2018. URL https://clinicaltrials.gov/ct2/show/NCT03087097. Accessed November 28, 2018.

[9] Samuel P. Costello, Oliver Waters, Robert V. Bryant, Rosa Katsikeros, Jesica Makanyanga, Mark Schoeman, Reme E. Mountifield, Derrick Tee, Stuart Howell, Patrick A. Hughes, Michael Conlon, Ian C. Roberts-Thomson, and Jane M. Andrews. Short duration, low intensity, pooled fecal microbiota transplantation induces remission in patients with mild-moderately active ulcerative colitis: A randomised controlled trial. *Gastroenterology*, 152(5):S198–S199, apr 2017. doi: 10.1016/s0016-5085(17)30969-1. URL https://doi.org/10.1016/s0016-5085(17)30969-1.

[10] Sudarshan Paramsothy, Michael A Kamm, Nadeem O Kaakoush, Alissa J Walsh, Johan van den Bogaerde, Douglas Samuel, Rupert W L Leong, Susan Connor, Watson Ng, Ramesh Paramsothy, Wei Xuan, Enmoore Lin, Hazel M Mitchell, and Thomas J Borody. Multidonor intensive faecal microbiota transplantation for active ulcerative colitis: a randomised placebo-controlled trial. *The Lancet*, 389(10075):1218–1228, mar 2017. doi: 10.1016/s0140-6736(17)30182-4. URL https://doi.org/10.1016/s0140-6736(17)30182-4.

[11] Stephan J. Ott, Georg H. Waetzig, Ateequr Rehman, Jacqueline Moltzau-Anderson, Richa Bharti, Juris A. Grasis, Liam Cassidy, Andreas Tholey, Helmut Fickenscher, Dirk Seegert, Philip Rosenstiel, and Stefan Schreiber. Efficacy of sterile fecal filtrate transfer for treating patients with clostridium difficile infection. *Gastroenterology*, 152(4):799–811.e7, mar 2017. doi: 10.1053/j.gastro.2016.11.010. URL https://doi.org/10.1053/j.gastro.2016.11.010.

[12] Tao Zuo, Sunny H Wong, Kelvin Lam, Rashid Lui, Kitty Cheung, Whitney Tang, Jessica Y L Ching, Paul K S Chan, Martin C W Chan, Justin C Y Wu, Francis K L Chan, Jun Yu, Joseph J Y Sung, and Siew C Ng. Bacteriophage transfer during faecal microbiota transplantation in clostridium difficile infection is associated with treatment outcome. *Gut*, pages 634–643, 2018. doi: 10.1136/gutjnl-2017-313952. URL https://doi.org/10.1136/gutjnl-2017-313952.

[13] Tanya Yatsunenko, Federico E. Rey, Mark J. Manary, Indi Trehan, Maria Gloria Dominguez-Bello, Monica Contreras, Magda Magris, Glida Hidalgo, Robert N. Baldassano, Andrey P. Anokhin, Andrew C. Heath, Barbara Warner, Jens Reeder, Justin Kuczynski, J. Gregory Caporaso, Catherine A. Lozupone, Christian Lauber, Jose Carlos Clemente, Dan Knights, Rob Knight, and Jeffrey I. Gordon. Human gut microbiome viewed across age and geography. *Nature*, 486(7402):222–227, may 2012. doi: 10.1038/nature11053. URL https://doi.org/10.1038/nature11053.

[14] AÃŕda Bafeta, Amelie Yavchitz, Carolina Riveros, Rui Batista, and Philippe Ravaud. Methods and reporting studies assessing fecal microbiota transplantation. *Annals of Internal Medicine*, 167(1):34, may 2017. doi: 10.7326/m16-2810. URL https://doi.org/10.7326/m16-2810.

[15] Scott W. Olesen, McKenzie M. Leier, Eric J. Alm, and Stacy A. Kahn. Searching for superstool: maximizing the therapeutic potential of FMT. *Nature Reviews Gastroenterology & Hepatology*, 15(7):387–388, apr 2018. doi: 10.1038/s41575-018-0019-4. URL https://doi.org/10.1038/s41575-018-0019-4.

[16] Paul Moayyedi, Michael G. Surette, Peter T. Kim, Josie Libertucci, Melanie Wolfe, Catherine Onischi, David Armstrong, John K. Marshall, Zain Kassam, Walter Reinisch, and Christine H. Lee. Fecal microbiota transplantation induces remission in patients with active ulcerative colitis in a randomized controlled trial. *Gastroenterology*, 149(1):102–109.e6, jul 2015. doi: 10.1053/j.gastro.2015.04.001. URL https://doi.org/10.1053/j.gastro.2015.04.001.

[17] Colleen R. Kelly, Alexander Khoruts, Christopher Staley, Michael J. Sadowsky, Mortadha Abd, Mustafa Alani, Brianna Bakow, Patrizia Curran, Joyce McKenney, Allison Tisch, Steven E. Reinert, Jason T. Machan, and Lawrence J. Brandt. Effect of fecal microbiota transplantation on recurrence in multiply recurrent clostridium difficile infection. *Annals of Internal Medicine*, 165(9):609, aug 2016. doi: 10.7326/m16-0271. URL https://doi.org/10.7326/m16-0271.

[18] Els van Nood, Anne Vrieze, Max Nieuwdorp, Susana Fuentes, Erwin G. Zoetendal, Willem M. de Vos, Caroline E. Visser, Ed J. Kuijper, Joep F.W.M. Bartelsman, Jan G.P. Tijssen, Peter Speelman, Marcel G.W. Dijkgraaf, and Josbert J. Keller. Duodenal infusion of donor feces for recurrent clostridium difficile. *New England Journal of Medicine*, 368(5):407–415, jan 2013. doi: 10.1056/nejmoa1205037. URL https://doi.org/10.1056/nejmoa1205037.

[19] Scott W Olesen, Thomas Gurry, and Eric J Alm. Designing fecal microbiota transplant trials that account for differences in donor stool efficacy. *Statistical Methods in Medical Research*, 27(10):2906–2917, feb 2017. doi: 10.1177/0962280216688502. URL https://doi.org/10.1177/0962280216688502.

[20] Robert A. Britton and Vincent B. Young. Role of the intestinal microbiota in resistance to colonization by clostridium difficile. *Gastroenterology*, 146(6):1547–1553, may 2014. doi: 10.1053/j.gastro.2014.01.059. URL https://doi.org/10.1053/j.gastro.2014.01.059.

[21] Jesse R. Zaneveld, Ryan McMinds, and Rebecca Vega Thurber. Stress and stability: applying the anna karenina principle to animal microbiomes. *Nature Microbiology*, 2(9):17121, aug 2017. doi: 10.1038/nmicrobiol.2017.121. URL https://doi.org/10.1038/nmicrobiol.2017.121.

[22] Majdi Osman, Zachery Stoltzner, Kelsey O'Brien, Kelly Ling, Emily Koelsch, Nancy Dubois, Kanchana Amaratunga, Mark Smith, and Zain Kassam. Donor efficacy in fecal microbiota transplantation for recurrent clostridium difficile: Evidence from a 1, 999-patient cohort. *Open Forum Infectious Diseases*, 3(suppl_1), 2016. doi: 10.1093/ofid/ofw194.48. URL https://doi.org/10.1093/ofid/ofw194.48.

[23] Ansel Hsiao, A. M. Shamsir Ahmed, Sathish Subramanian, Nicholas W. Griffin, Lisa L. Drewry, William A. Petri, Rashidul Haque, Tahmeed Ahmed, and Jeffrey I. Gordon. Members of the human gut microbiota involved in recovery from vibrio cholerae infection. *Nature*, 515(7527):423–426, sep 2014. doi: 10.1038/nature13738. URL https://doi.org/10.1038/nature13738.

[24] Nicola Wilck, Mariana G. Matus, Sean M. Kearney, Scott W. Olesen, Kristoffer Forslund, Hendrik Bartolomaeus, Stefanie Haase, Anja Mähler, András Balogh, Lajos Markó, Olga Vvedenskaya, Friedrich H. Kleiner, Dmitry Tsvetkov, Lars Klug, Paul I. Costea, Shinichi Sunagawa, Lisa Maier, Natalia Rakova, Valentin Schatz, Patrick Neubert, Christian Frätzer, Alexander Krannich, Maik Gollasch, Diana A. Grohme, Beatriz F. Côrte-Real, Roman G. Gerlach, Marijana Basic, Athanasios Typas, Chuan Wu, Jens M. Titze, Jonathan Jantsch, Michael Boschmann, Ralf Dechend, Markus Kleinewietfeld, Stefan Kempa, Peer Bork, Ralf A. Linker, Eric J. Alm, and Dominik N. Müller. Salt-responsive gut commensal modulates TH17 axis and disease. *Nature*, nov 2017. doi: 10.1038/nature24628. URL https://doi.org/10.1038/nature24628.

[25] Aleksandar D. Kostic, Eunyoung Chun, Lauren Robertson, Jonathan N. Glickman, Carey Ann Gallini, Monia Michaud, Thomas E. Clancy, Daniel C. Chung, Paul Lochhead, Georgina L. Hold, Emad M. El-Omar, Dean Brenner, Charles S. Fuchs, Matthew Meyerson, and Wendy S. Garrett. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host & Microbe*, 14(2):207–215, aug 2013. doi: 10.1016/j.chom.2013.07.007. URL https://doi.org/10.1016/j.chom.2013.07.007.

[26] Mara Roxana Rubinstein, Xiaowei Wang, Wendy Liu, Yujun Hao, Guifang Cai, and Yiping W. Han. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating e-cadherin/β-catenin signaling via its FadA adhesin. *Cell Host & Microbe*, 14(2):195–206, aug 2013. doi: 10.1016/j.chom.2013.07.012. URL `https://doi.org/10.1016/j.chom.2013.07.012`.

[27] Susan Bullman, Chandra S. Pedamallu, Ewa Sicinska, Thomas E. Clancy, Xiaoyang Zhang, Diana Cai, Donna Neuberg, Katherine Huang, Fatima Guevara, Timothy Nelson, Otari Chipashvili, Timothy Hagan, Mark Walker, Aruna Ramachandran, Begoña Diosdado, Garazi Serna, Nuria Mulet, Stefania Landolfi, Santiago Ramon y Cajal, Roberta Fasani, Andrew J. Aguirre, Kimmie Ng, Elena Élez, Shuji Ogino, Josep Tabernero, Charles S. Fuchs, William C. Hahn, Paolo Nuciforo, and Matthew Meyerson. Analysis ofFusobacteriumpersistence and antibiotic response in colorectal cancer. *Science*, 358(6369):1443–1448, nov 2017. doi: 10.1126/science.aal5240. URL `https://doi.org/10.1126/science.aal5240`.

[28] Effect of butyrate enemas on the colonic mucosa in distal ulcerative colitis. *Gastroenterology*, 103:51–56, 1992.

[29] Claire Duvallet, Sean M. Gibbons, Thomas Gurry, Rafael A. Irizarry, and Eric J. Alm. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nature Communications*, 8(1), dec 2017. doi: 10.1038/s41467-017-01973-8. URL `https://doi.org/10.1038/s41467-017-01973-8`.

[30] Melanie Schirmer, Lee Denson, Hera Vlamakis, Eric A. Franzosa, Sonia Thomas, Nathan M. Gotman, Paul Rufo, Susan S. Baker, Cary Sauer, James Markowitz, Marian Pfefferkorn, Maria Oliva-Hemker, Joel Rosh, Anthony Otley, Brendan Boyle, David Mack, Robert Baldassano, David Keljo, Neal LeLeiko, Melvin Heyman, Anne Griffiths, Ashish S. Patel, Joshua Noe, Subra Kugathasan, Thomas Walters, Curtis Huttenhower, Jeffrey Hyams, and Ramnik J. Xavier. Compositional and temporal changes in the gut microbiome of pediatric ulcerative colitis patients are linked to disease course. *Cell Host & Microbe*, 24(4):600–610.e4, oct 2018. doi: 10.1016/j.chom.2018.09.009. URL `https://doi.org/10.1016/j.chom.2018.09.009`.

[31] P. Kump, P. Wurm, H. P. Gröchenig, H. Wenzl, W. Petritsch, B. Halwachs, M. Wagner, V. Stadlbauer, A. Eherer, K. M. Hoffmann, A. Deutschmann, G. Reicht, L. Reiter, P. Slawitsch, G. Gorkiewicz, and C. Högenauer. The taxonomic composition of the donor intestinal microbiota is a major factor influencing the efficacy of faecal microbiota transplantation in therapy refractory ulcerative colitis. *Alimentary Pharmacology & Therapeutics*, 47(1):67–77, jan 2018. doi: 10.1111/apt.14387. URL `https://doi.org/10.1111/apt.14387`.

[32] Alka Goyal, Andrew Yeh, Brian R Bush, Brian A Firek, Leah M Siebold, Matthew Brian Rogers, Adam D Kufen, and Michael J Morowitz. Safety, clinical response, and microbiome findings following fecal microbiota transplant in

children with inflammatory bowel disease. *Inflammatory Bowel Diseases*, 24(2): 410–421, jan 2018. doi: 10.1093/ibd/izx035. URL https://doi.org/10.1093/ibd/izx035.

[33] Vinita Jacob, Carl Crawford, Shirley Cohen-Mekelburg, Monica Viladomiu, Gregory G. Putzel, Yecheskel Schneider, Fatiha Chabouni, Sarah O'Neil, Brian Bosworth, Viola Woo, Nadim J. Ajami, Joseph F. Petrosino, Ylaine Gerardin, Zain Kassam, Mark Smith, Iliyan D. Iliev, Gregory F. Sonnenberg, David Artis, Ellen Scherl, and Randy S. Longman. Single delivery of high-diversity fecal microbiota preparation by colonoscopy is safe and effective in increasing microbial diversity in active ulcerative colitis. *Inflammatory Bowel Diseases*, 23(6):903–911, jun 2017. doi: 10.1097/mib.0000000000001132. URL https://doi.org/10.1097/mib.0000000000001132.

[34] Marius Vital, André Karch, and Dietmar H. Pieper. Colonic butyrate-producing communities in humans: an overview using omics data. *mSystems*, 2(6), dec 2017. doi: 10.1128/msystems.00130-17. URL https://doi.org/10.1128/msystems.00130-17.

[35] Jasmohan S. Bajaj, Zain Kassam, Andrew Fagan, Edith A. Gavis, Eric Liu, I. Jane Cox, Raffi Kheradman, Douglas Heuman, Jessica Wang, Thomas Gurry, Roger Williams, Masoumeh Sikaroodi, Michael Fuchs, Eric Alm, Binu John, Leroy R. Thacker, Antonio Riva, Mark Smith, Simon D. Taylor-Robinson, and Patrick M Gillevet. Fecal microbiota transplant from a rational stool donor improves hepatic encephalopathy: A randomized clinical trial. *Hepatology*, 66 (6):1727–1738, oct 2017. doi: 10.1002/hep.29306. URL https://doi.org/10.1002/hep.29306.

[36] Alan F. Hofmann. Enterohepatic circulation of bile acids. *Handbook of Physiology. The Gastrointestinal System.*, 4:567–596, 1989. doi: 10.1002/cphy.cp060329. URL https://doi.org/10.1002/cphy.cp060329.

[37] X. Wang and G.R. Gibson. Effects of thein vitrofermentation of oligofructose and inulin by bacteria growing in the human large intestine. *Journal of Applied Bacteriology*, 75(4):373–380, oct 1993. doi: 10.1111/j.1365-2672.1993.tb02790.x. URL https://doi.org/10.1111/j.1365-2672.1993.tb02790.x.

[38] Tingting Chen, Wenmin Long, Chenhong Zhang, Shuang Liu, Liping Zhao, and Bruce R. Hamaker. Fiber-utilizing capacity varies in prevotella- versus bacteroides-dominated gut microbiota. *Scientific Reports*, 7(1), jun 2017. doi: 10.1038/s41598-017-02995-4. URL https://doi.org/10.1038/s41598-017-02995-4.

[39] Robert A Koeth, Zeneng Wang, Bruce S Levison, Jennifer A Buffa, Elin Org, Brendan T Sheehy, Earl B Britt, Xiaoming Fu, Yuping Wu, Lin Li, Jonathan D Smith, Joseph A DiDonato, Jun Chen, Hongzhe Li, Gary D Wu, James D Lewis, Manya Warrier, J Mark Brown, Ronald M Krauss, W H Wilson Tang,

Frederic D Bushman, Aldons J Lusis, and Stanley L Hazen. Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nature Medicine*, 19(5):576–585, apr 2013. doi: 10.1038/nm.3145. URL https://doi.org/10.1038/nm.3145.

[40] Zeneng Wang, Adam B. Roberts, Jennifer A. Buffa, Bruce S. Levison, Weifei Zhu, Elin Org, Xiaodong Gu, Ying Huang, Maryam Zamanian-Daryoush, Miranda K. Culley, Anthony J. DiDonato, Xiaoming Fu, Jennie E. Hazen, Daniel Krajcik, Joseph A. DiDonato, Aldons J. Lusis, and Stanley L. Hazen. Non-lethal inhibition of gut microbial trimethylamine production for the treatment of atherosclerosis. *Cell*, 163(7):1585–1595, dec 2015. doi: 10.1016/j.cell.2015.11.055. URL https://doi.org/10.1016/j.cell.2015.11.055.

[41] L. V. Blanton, M. R. Charbonneau, T. Salih, M. J. Barratt, S. Venkatesh, O. Ilkaveya, S. Subramanian, M. J. Manary, I. Trehan, J. M. Jorgensen, Y. m. Fan, B. Henrissat, S. A. Leyn, D. A. Rodionov, A. L. Osterman, K. M. Maleta, C. B. Newgard, P. Ashorn, K. G. Dewey, and J. I. Gordon. Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. *Science*, 351(6275):aad3311–aad3311, feb 2016. doi: 10.1126/science.aad3311. URL https://doi.org/10.1126/science.aad3311.

[42] Sathish Subramanian, Sayeeda Huq, Tanya Yatsunenko, Rashidul Haque, Mustafa Mahfuz, Mohammed A. Alam, Amber Benezra, Joseph DeStefano, Martin F. Meier, Brian D. Muegge, Michael J. Barratt, Laura G. VanArendonk, Qunyuan Zhang, Michael A. Province, William A. Petri Jr, Tahmeed Ahmed, and Jeffrey I. Gordon. Persistent gut microbiota immaturity in malnourished bangladeshi children. *Nature*, 510(7505):417–421, jun 2014. doi: 10.1038/nature13421. URL https://doi.org/10.1038/nature13421.

[43] Jakob Stokholm, Martin J. Blaser, Jonathan Thorsen, Morten A. Rasmussen, Johannes Waage, Rebecca K. Vinding, Ann-Marie M. Schoos, Asja Kunøe, Nadia R. Fink, Bo L. Chawes, Klaus Bønnelykke, Asker D. Brejnrod, Martin S. Mortensen, Waleed Abu Al-Soud, Søren J. Sørensen, and Hans Bisgaard. Maturation of the gut microbiome and risk of asthma in childhood. *Nature Communications*, 9(1), jan 2018. doi: 10.1038/s41467-017-02573-2. URL https://doi.org/10.1038/s41467-017-02573-2.

[44] Laura M. Cox, Shingo Yamanishi, Jiho Sohn, Alexander V. Alekseyenko, Jacqueline M. Leung, Ilseung Cho, Sungheon G. Kim, Huilin Li, Zhan Gao, Douglas Mahana, Jorge G. Zárate Rodriguez, Arlin B. Rogers, Nicolas Robine, P'ng Loke, and Martin J. Blaser. Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. *Cell*, 158(4):705–721, aug 2014. doi: 10.1016/j.cell.2014.05.052. URL https://doi.org/10.1016/j.cell.2014.05.052.

[45] Aleksandar D. Kostic, Dirk Gevers, Heli Siljander, Tommi Vatanen, Tuulia Hyötyläinen, Anu-Maaria Hämäläinen, Aleksandr Peet, Vallo Tillmann, Päivi

Pöhö, Ismo Mattila, Harri Lähdesmäki, Eric A. Franzosa, Outi Vaarala, Marcus de Goffau, Hermie Harmsen, Jorma Ilonen, Suvi M. Virtanen, Clary B. Clish, Matej Orešič, Curtis Huttenhower, Mikael Knip, and Ramnik J. Xavier. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host & Microbe*, 17(2):260–273, feb 2015. doi: 10.1016/j.chom.2015.01.001. URL https://doi.org/10.1016/j.chom.2015.01.001.

[46] Richard R Stein, Takeshi Tanoue, Rose L Szabady, Shakti K Bhattarai, Bernat Olle, Jason M Norman, Wataru Suda, Kenshiro Oshima, Masahira Hattori, Georg K Gerber, Chris Sander, Kenya Honda, and Vanni Bucci. Computer-guided design of optimal microbial consortia for immune system modulation. *eLife*, 7, apr 2018. doi: 10.7554/elife.30916. URL https://doi.org/10.7554/elife.30916.

[47] Jonas Halfvarson, Colin J. Brislawn, Regina Lamendella, Yoshiki Vazquez-Baeza, William A. Walters, Lisa M. Bramer, Mauro D'Amato, Ferdinando Bonfiglio, Daniel McDonald, Antonio Gonzalez, Erin E. McClure, Mitchell F. Dunklebarger, Rob Knight, and Janet K. Jansson. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nature Microbiology*, 2(5), feb 2017. doi: 10.1038/nmicrobiol.2017.4. URL https://doi.org/10.1038/nmicrobiol.2017.4.

[48] Ashley Shade and Jo Handelsman. Beyond the venn diagram: the hunt for a core microbiome. *Environmental Microbiology*, 14(1):4–12, oct 2011. doi: 10.1111/j.1462-2920.2011.02585.x. URL https://doi.org/10.1111/j.1462-2920.2011.02585.x.

[49] A. M. Schubert, M. A. M. Rogers, C. Ring, J. Mogle, J. P. Petrosino, V. B. Young, D. M. Aronoff, and P. D. Schloss. Microbiome data distinguish patients with clostridium difficile infection and non-c. difficile-associated diarrhea from healthy controls. *mBio*, 5(3), may 2014. doi: 10.1128/mbio.01021-14. URL https://doi.org/10.1128/mbio.01021-14.

[50] Nielson T. Baxter, Mack T. Ruffin, Mary A. M. Rogers, and Patrick D. Schloss. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Medicine*, 8(1), apr 2016. doi: 10.1186/s13073-016-0290-3. URL https://doi.org/10.1186/s13073-016-0290-3.

[51] Julia K. Goodrich, Jillian L. Waters, Angela C. Poole, Jessica L. Sutter, Omry Koren, Ran Blekhman, Michelle Beaumont, William Van Treuren, Rob Knight, Jordana T. Bell, Timothy D. Spector, Andrew G. Clark, and Ruth E. Ley. Human genetics shape the gut microbiome. *Cell*, 159(4):789–799, nov 2014. doi: 10.1016/j.cell.2014.09.053. URL https://doi.org/10.1016/j.cell.2014.09.053.

[52] Junhai Ou, Franck Carbonero, Erwin G Zoetendal, James P DeLany, Mei Wang, Keith Newton, H Rex Gaskins, and Stephen JD O'Keefe. Diet, microbiota, and microbial metabolites in colon cancer risk in rural africans and african americans. *The American Journal of Clinical Nutrition*, 98(1):111–120, may 2013. doi: 10.3945/ajcn.112.056689. URL https://doi.org/10.3945/ajcn.112.056689.

[53] C. De Filippo, D. Cavalieri, M. Di Paola, M. Ramazzotti, J. B. Poullet, S. Massart, S. Collini, G. Pieraccini, and P. Lionetti. Impact of diet in shaping gut microbiota revealed by a comparative study in children from europe and rural africa. *Proceedings of the National Academy of Sciences*, 107(33):14691–14696, aug 2010. doi: 10.1073/pnas.1005963107. URL https://doi.org/10.1073/pnas.1005963107.

[54] Maria G. Dominguez Bello, Rob Knight, Jack A. Gilbert, and Martin J. Blaser. Preserving microbial diversity. *Science*, 362(6410):33–34, oct 2018. doi: 10.1126/science.aau8816. URL https://doi.org/10.1126/science.aau8816.

[55] Tania Rabesandratana. 'poop vault' of human feces could preserve gut's microbial biodiversity—and help treat disease. *Science*, nov 2018. doi: 10.1126/science.aav9276. URL https://doi.org/10.1126/science.aav9276.

[56] Bolyen E, Rideout JR, Dillon MR, Bokulich NA, and et al. Qiime 2: Reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ Preprints*, 6(e27295v2), 2018. doi: 10.1038/nmeth.f.303.

[57] Amnon Amir, Daniel McDonald, Jose A Navas-Molina, Evguenia Kopylova, James T Morton, Zhenjiang Zech Xu, Eric P Kightley, Luke R Thompson, Embriette R Hyde, Antonio Gonzalez, and Rob Knight. Deblur rapidly resolves single-nucleotide community sequence patterns. *MSystems*, 2(2):e00191–16, 2017.

[58] Nicholas A. Bokulich, Benjamin D. Kaehler, Jai Ram Rideout, Matthew Dillon, Evan Bolyen, Rob Knight, Gavin A. Huttley, and J. Gregory Caporaso. Optimizing taxonomic classification of marker-gene amplicon sequences with qiime 2's q2-feature-classifier plugin. *Microbiome*, 6(1):90, 2018. doi: 10.1186/s40168-018-0470-z. URL https://doi.org/10.1186/s40168-018-0470-z.

[59] Genta Kakiyama, William M. Pandak, Patrick M. Gillevet, Phillip B. Hylemon, Douglas M. Heuman, Kalyani Daita, Hajime Takei, Akina Muto, Hiroshi Nittono, Jason M. Ridlon, Melanie B. White, Nicole A. Noble, Pamela Monteith, Michael Fuchs, Leroy R. Thacker, Masoumeh Sikaroodi, and Jasmohan S. Bajaj. Modulation of the fecal bile acid profile by gut microbiota in cirrhosis. *Journal of Hepatology*, 58(5):949–955, may 2013. doi: 10.1016/j.jhep.2013.01.003. URL https://doi.org/10.1016/j.jhep.2013.01.003.

Chapter 6

Residential sewage provides a view into community-level human microbiome and metabolome

Mariana Matus, Claire Duvallet, Melissa Kido Soule, Krista Longnecker, Noriko Endo, Newsha Ghaeli, Ilana Brito, Carlo Ratti, Elizabeth B Kujawinski, Eric J Alm

The contents of this chapter constitute a manuscript in preparation for submission. The main figures are at the end of the chapter and the supplementary figures are in Appendix C.

Abstract

Wastewater-based epidemiology has long been proposed as a tool for population health monitoring, with applications ranging from infectious disease surveillance to drug consumption estimation. To date, however, there are few examples of wastewater epidemiology being implemented in practice. Here, we propose a novel platform that adapts existing wastewater-based epidemiology approaches for community-level public health surveillance. We integrate sewage network and population information to identify a sampling manhole representative of a residential community and show that the wastewater sampled at this upstream manhole contains more human-derived bacteria and metabolites than downstream wastewater treatment plant samples, which are the usual sampling locations in wastewater epidemiology. Our method to perform untargeted metabolomics on raw sewage allows for the identification of multiple glucuronide compounds, which are indicative of direct human excretion and typically too unstable to be detected in wastewater. We show that these compounds are absent in downstream samples, but abundant in our residential sewage samples. Next, we perform an hourly 24-hour sampling campaign to identify human and environmental components of the sewage metabolome and microbiome based on diurnal variations. We confirm urinary and fecal metabolites and find that their dynamics reflect human behavior but differ from each other. Finally, we propose that mining untargeted data derived from residential sewage opens additional possibilities for identifying biomarkers with direct public health or policy relevance and show preliminary results from the metabolomics data. Together, these results suggest a novel approach for implementing wastewater epidemiology in urban settings that provides a new source of community-level health data and which could be developed into a citywide public health observatory.

6.1 Introduction

Urbanization is a globally increasing phenomenon. By 2050, 70% of the world's population will live in cities where 80% of the global GDP will be generated. Cities are thus prime locations to deliver services that improve the wellbeing of residents. However, public health officials in cities tend to operate under financial constraints and with limited data to establish priorities and measure the impact of their programs. Wastewater has been proposed as a source of data on population health and behavior that can fill this gap at relatively low cost for public health officials [1].

Wastewater-based epidemiology could have enormous benefit to public health practice, providing real-time assessment of population health and generating data to support public health and city policies. Currently, most public health surveillance is based on reports from hospitals and other clinical centers, and appropriate responses are initiated after enough clinical cases have been reported. Wastewater-based public health surveillance platforms, in contrast, could provide real-time data on emerging outbreaks before a critical mass of patients have gotten sick [2, 3]. Viral and bacterial pathogens shed in stool could be directly detected in sewage, triggering the deployment of coordinated awareness campaigns or other public health interventions to areas with the largest number of sick patients, all without needing to wait for case reports from hospitals. In addition to identifying emerging outbreaks, wastewater could provide a novel source of data to evaluate the effectiveness of policies intended to have long-term effects on a population's health [1]. For example, rather than waiting years to measure the impact of taxing sugary drinks on population obesity levels, biomarkers of obesity or consumption could be directly measured in sewage before and after the policy roll-out, thus providing a more immediate measure of the impact of policy on targeted outcomes.

Wastewater-based epidemiology has been proposed for many exciting applications. It has been shown to be useful as an early indicator of poliovirus outbreaks to target vaccination campaigns in polio-free countries [2, 4] and to quantify asymptomatic circulation [5]. Environmental surveillance of typhoid and antimicrobial resistance

through urban wastewater are emerging areas of interest in global and public health organizations [6, 7]. Besides infectious disease, methods have been developed and standardized over the last decade to successfully provide continuous data on illicit drug consumption at the national- and regional-level in Australia [8] and over 60 European cities and towns [9, 10, 11]. The academic community is now interested in branching out to measure lifestyle chemicals and biomarkers of health such as obesity levels [1, 12, 13]. More recently, wastewater testing has been proposed as a promising tool to tackle the opioid epidemic that is severely affecting the U.S [14, 15]. However, apart from successful poliovirus surveillance and, to a lesser extent, national-level drug consumption surveillance programs, most application areas remain proof-of-concept academic projects and have yet to be implemented into public health and other public service delivery agencies.

Despite great technical advances over the past decades, wastewater testing still presents considerable challenges that decrease its usefulness to public health officials and policy-makers. In most proposed applications, wastewater is sampled at downstream sites like pump stations or treatment plants [2, 3]. While downstream collection sites are easily accessible and cost-effective, they impose barriers to data interpretation in the public health context. Wastewater at treatment plants is a combination of residential households and commercial and industrial buildings, and it may represent one or more municipalities. These factors create technical biases in the data and prevent clear consumption or infection rates from being calculated from measured quantities. Additionally, if the interest is to compare several cities, collection from different wastewater treatment plants or other downstream sites may introduce confounding factors, such as representation of different population sizes [16], large variation in wastewater travel time from different sources, different rates of in-sewer degradation of chemical [17] and bacterial biomarkers, and collection with ad-hoc sampling equipment that prevents researchers from optimizing collection parameters to obtain representative samples [18]. Finally, while collecting composite samples from wastewater treatment plants (WWTPs) is practical and cost-effective to generate national estimates [19, 20, 10], data from WWTPs is not optimal for city

officials since it lacks geographic specificity. Finally, public health officials are interested in a broad range of human health biomarkers but it is not straightforward to know what markers are stable and quantifiable in wastewater collected from different regions [12].

Here, we propose a novel implementation of wastewater epidemiology that addresses these challenges and develops wastewater testing as a platform to directly measure the health of urban communities and facilitate applications to public health. We combine demographic, sewage, and geographic data sources to curate and select an upstream sampling site that reflects a residential population. By mining untargeted microbiome and metabolomics data, we show that residential sewage reflects human diurnal activity and identify human-derived biomarkers which can be specifically separated from environmental contributors. Together, this work is the first to show that upstream wastewater sampling could provide a novel useful data source for public health agencies.

6.2   Results

6.2.1   Upstream residential catchments provide more useful public health information than WWTP samples

Integrating GIS datasets with demographic and sewage network information enables the selection of sampling sites with carefully curated technical and population characteristics. We defined residential catchments as those which represent areas with at least 80% residential land use, a population of over 5,000 people, and a maximum sewage travel time of less than 60 minutes (Fig. 6-1A). We screened all manholes in our municipality and selected a manhole which met these criteria (Fig. 6-1A). From a usability perspective, data from residential catchments will be more actionable by municipal public health departments because sites can be selected to mostly capture resident rather than transient populations, and to have the spatial resolution that matches their public health interventions.

Besides providing geographical specificity to the sample, residential catchments mitigate variability in population size and sewage travel time observed at WWTPs. Our selection criteria produces catchments with similar population sizes by design (5,000–10,000 inhabitants) rather than the large range observed (1,000–2,000,000 inhabitants) in WWTPs (Supp. Fig. C-1, [13]). Additionally, residential catchments are more likely to contain even representation of all households in the catchment compared to WWTPs: in our residential catchment, sewage from every source travels at most 45 minutes to the sampling point and the distribution of travel times is narrowly-distributed (Fig. 6-1A). Sewage at WWTPs, on the other hand, can have travelled anywhere from minutes to hours before it reaches the sampling point. Sampling in residential catchments therefore mitigates the need to correct for unequal population sizes and travel times to some extent.

Residential sewage contains more direct chemical markers of human consumption and behavior than sewage sampled at WWTPs. We sampled a residential sewer (selected as described above) and a downstream pump station (Fig. 6-1B) and performed untargeted metabolomics and 16S rRNA sequencing on the filtrate and bacterial cells, respectively. 34.5% of the metabolites detected in both sewer and WWTP samples (n=245/710) had significantly different concentrations (p-value < 0.05 independent T-test with FDR correction for multiple testing). Of these differentially abundant metabolites, 73% (n=179/245) decreased or disappeared completely from wastewater collected at the WWTP, including many metabolites from human activity as defined later in this paper (Fig. 6-2A yellow dots, Fig. 6-3B yellow bars, Supp. Fig. C-2 cluster M1). Similarly, we found that residential sewage contained a higher proportion of human-associated bacteria than WWTP sewage. On average, 66% of the microbial community from residential sewage was of human fecal origin, compared to 51% in the pump station samples (Fig. 6-2B), with the Ruminococcaceae, Lachnospiraceae, Prevotellaceae, Veillonellaceae and Erysipelotrichaceae families being reduced (Supp. Fig. C-3). These results are similar to those reported in Newton et al. [13], where on average 15% of sequences in WWTP sewage had human fecal origin. Together, these results indicate that wastewater from residential catchments which has been traveling

149

in sewers for less than one hour contains many human biomarkers that are absent in wastewater collected at WWTPs.

Furthermore, performing untargeted metabolomics on residential catchment samples extends the possible biomarkers that can be monitored through wastewater epidemiology. We identified 22 glucuronide compounds in our untargeted data (see Methods). Glucuronides are chemical groups added to exogenous molecules by the human liver to aid in their excretion and are thus indicators of human excretion, distinguishing between molecules simply discarded into the sewers versus those consumed and excreted by humans and providing direct evidence of consumption. However, many bacteria carry glucuronidase enzymes that release the glucuronide group to use as carbon source [21], and these unstable molecules have previously been difficult to identify in sewage samples from WWTPs [22]. In our data, the majority of the identified glucuronides were significantly reduced or altogether absent in wastewater collected at the WWTP (Figure 6-2C). Sampling upstream allows for the detection of these unstable molecules and also more generally expands the number of compounds which can be measured. Molecules which are usually difficult to measure by mass spectrometry in wastewater because they do not ionize well should be captured in residential catchments, since the human body excretes compounds by dissolving them in urine or stool and therefore naturally converts exogenous compounds into more ionizable chemical forms. Thus, detecting a large number of glucuronides in our residential catchment sewage sample suggests that expanding wastewater monitoring to chemical variants of molecules of interest may be beneficial for future environmental surveillance applications.

6.2.2 Residential sewage microbiome and metabolome contain both human- and environmentally-derived components

Chemical and bacterial biomarkers detected in residential sewage reflect the daily timescale of human activity. We used the average sewage flow rate over two months as a proxy for human activity (Figure 6-3). We hypothesized that human-associated

metabolites and bacteria would increase during the day, when the contributing population would be more active, and decrease at night. We collected hourly samples over 24 hours from the residential catchment manhole and produced 16S rDNA sequencing and untargeted metabolomics profiles (Figure 6-3A). We co-clustered metabolomic features (n=3,672) and microbial OTUs (n=254) based on their Spearman correlation (Fig. 6-3B & D, Supp. Fig. C-2), and found that the metabolic features grouped into two main clusters (M1 and M2), while bacterial OTUs produced three groups (O1 and O2 are highlighted). The M1 metabolite cluster (n=2,815) had strong positive correlation with the O1 bacterial cluster (n=41) and negative correlation to the O2 bacterial cluster (n=84). The majority of metabolites and bacteria in these groups also correlated with the average flow rate in the sewer, increasing during the day and becoming drastically less abundant at night (Fig. 6-3B & D).

Metabolites and bacteria in sewage separate into human and environmental contributions. Groups M1 and O1 were enriched in human-derived bacteria and metabolites, while groups M2 and O2 contained bacteria and metabolites derived from the environment (Fig. 6-3C). The cluster M1 was statistically enriched in matches to human metabolites compared to the M2 cluster (Fig. 6-3C), and includes known urinary and fecal metabolites. Twenty-two metabolic features in M1 were identified as glucuronide compounds of hormones, bile acids and pharmaceuticals, which are direct markers of human excretion as discussed above. Bacteria in the cluster O1 were putatively identified as coming from human gut microbiomes. Cluster O1 included 41 bacterial OTUs primarily from the Firmicutes and Bacteroidetes phyla, which are major components of the human gut flora (Fig. 6-3D). Additionally, the top 10 most abundant bacterial families in human stool (Human Microbiome Project) made up 65% of the community in the O1 group (Supp. Fig. C-4). Similar to what was observed in the metabolites, O1 bacteria had a dip in relative abundance during nighttime (Figure 6-3B). Clusters M2 and O2, on the other hand, remained relatively constant through the day and likely reflect bacteria and metabolites sourced from environmental background rather than human activity. Cluster M2 had few putative human-associated hits, and had constant abundance through the day and nighttime.

The molecules in this group likely represent chemicals derived from the environment or sewer biochemistry. Human-associated bacteria were similarly decreased in cluster O2: most of the OTUs were Proteobacteria, which are not major members of the normal gut flora, and human fecal bacterial families comprised only 1.4% of the abundance of this cluster (Supp. Fig. C-4). The relative abundance of cluster O2 increased at night, reflecting that these environmental made up a larger proportion of the residential sewage microbiome during periods of low human activity (Fig. 6-3D).

6.2.3 Identification of potential public health biomarkers in residential sewage

Urine and fecal contributions to residential sewage can be identified through untargeted analyses. We identified and confirmed metabolites corresponding to human urinary and fecal markers through a combination of MS2 matching and confirming with analytical standards. In our attempts to annotate metabolite features, we prioritized features with high abundance across all samples and in a few samples of interest (i.e. 8 am and 2 pm samples, which contained more metabolites and which we expected to correlate with human behavior). We screened the top mz values against METLIN and HMDB databases to find putative human-derived metabolites [23, 24]. We confirmed features which putatively matched urinary or fecal markers and for which we could acquire analytical standards. Furthermore, we compared the MS2 fragmentation patterns for potential glucuronide compounds with their expected fragmentations and used diagnostic mz peaks to putatively confirm these compounds (see Methods). The most abundant feature in our dataset was confirmed via analytical standard as alpha-N-phenylacetyl-L-glutamine (PAG), a metabolite of phenylacetic acid and an end product of phenylalanine metabolism excreted in urine [25, 26]. Previous studies on human physiology have found that PAG excretion remains constant in response to diet and does not exhibit diurnal cycling in individual patients [25]. Thus, the abundance of PAG in our sample likely correlates with the total volume of urine in sewage at any given time. Urobilinogen, a breakdown product of hemoglobin excreted

primarily in stool [27], was also confirmed via analytical standard. We also confirmed 4 bile acids (chenodeoxycholic acid, cholic acid, glycocholic acid, taurocholic acid), which represent the major primary conjugated and unconjugated bile acids produced by the liver to aid in digestion and which we considered as putative fecal markers.

The urinary and fecal metabolites reflected general diurnal human dynamics, but differed from each other (Figure 6-4). Both types of metabolites decreased at night, with the fecal markers almost completely disappearing while the urinary marker remained detectable. Fecal markers also showed significant spikes in the 8 am and 3 pm samples, and were slightly elevated throughout the evening (8 - 10 pm). The urinary marker, on the other hand, increased in the morning and stayed relatively consistent throughout the day. These patterns correspond with known patterns of human behavior: at night, people are less likely to wake up to defecate but do wake up to urinate. During the day, urination remains relatively constant across the entire population, but defecation probably spikes in the morning right after waking up and after meal times.

Furthermore, the daily dynamics of bile acids in residential sewage may reflect population-level behavior and physiology. During the day, levels of unconjugated cholic and chenodeoxycholic acids remained relatively constant but conjugated cholic acids (glycocholic and taurocholic acid) spiked after mealtimes (Fig. 6-4). In individuals, the liver synthesizes more bile acids from cholesterol after it is ingested during a meal [28]. Levels of taurocholic acid are also known to increase significantly after meals, since the diet is the primary source of taurine in humans. The liver secretes only the conjugated forms of cholic acid, glycocholic and taurocholic acid, into the gut. Thus, the post-meal spikes in conjugated bile acids in our data may reflect not only more stool in the sewage but also an aggregate population-level increase in physiological production of bile acids by the liver in response to meals. The daily dynamics of fecal bile acid excretion has not been studied in humans, in part because it is difficult to access finely temporally-resolved data given that humans defecate only 1-3 times per day on average. These data suggest that analyzing physically aggregated biological data could provide another avenue by which to study human physiology or

measure its functioning.

However, an alternative explanation for these spikes in conjugated bile acids may be that they reflect only incidental increases in the sampling of individuals with diarrhea at these times. Diarrhea is characterized by a decrease in the conversion of conjugated primary bile acids to their unconjugated forms. Furthermore, because our metabolomics data is generated from the liquid phase of wastewater after removing solids, diarrhea will contribute relatively more than solid stools to the sampled metabolites at any given time. Thus, our results are consistent with a sampling enriched in individuals with diarrhea at times when they defecate. By contrast, the contribution of stools from healthy individuals (as indicated by the levels of cholate and deoxycholate) remains relatively constant over the course of the day. Taken together, the findings that fecal and urinary metabolites differ in their dynamics and that individual people may disproportionately contribute to individual samples, suggest that developing methods to correct for biases in sampling and normalize for the number of people contributing to a given sample will be required before these data can be acted upon in public health contexts.

6.3 Discussion

The platform that we present is the first to adapt the practices of wastewater-based epidemiology specifically for the needs of city-level public officials, and is the first to leverage untargeted multi-omics data to examine the breadth of human biomarkers available in residential wastewater. Through high-throughput microbiome and metabolome analyses, we show that wastewater from residential catchments is a closer reflection of human health than wastewater collected at downstream sites. Wastewater from residential catchments contains more human metabolites and human-associated bacteria than wastewater from downstream sites, including unstable glucuronides which can be used as direct markers of human consumption. Untargeted analyses can identify glucuronides and potentially other chemical variants of molecules, expanding the scope of compounds amenable to wastewater surveillance.

154

Moreover, metabolites and bacteria in residential sewage can be compartmentalized into human and environmental contributions, and human-associated components reflect known human behavior and activity. Finally, our data suggests that identifying biomarkers across many realms of interest such as human behavior, consumption, and health, may be possible through residential wastewater surveillance. However, incorporating residential sewage monitoring into public health departments will require addressing additional analytical, engineering, logistical, and ethical challenges.

First, models to estimate the population size contributing to a given sample will need to be developed in order to convert measured quantities into per capita rates. Samples taken from wastewater treatment plants can be reasonably expected to represent a large portion of the catchment population, representing thousands to hundreds of thousands of individuals. In contrast, the number of people represented in a residential sewage sample may fluctuate significantly in these smaller residential catchments. Without knowing how many people contributed to a given sample, it is not possible to interpret the quantities measured in a given sample: did one person excrete a large amount of the measured component, or did many people excrete a small amount of it? Knowing the"denominator" of what is measured in sewage will be required to translate measured quantities into values that city officials can act on.

Scaling this work to the city-level will also require the development of new sampling equipment sampling campaign designs. Sampling equipment could be developed to aggregate flow-proportional samples over a full day, and more complex samplers could also perform some pre-processing or basic measurements *in situ*. The sampling sites chosen in a given campaign could be optimized for a variety of purpose: covering as much of the population with the least number of sites, identifying hot spots where potential outbreaks are likely to begin, comparing neighborhoods with different socioeconomic characteristics or health behaviors, or standardizing sampling site characteristics to enable multi-city comparisons. Each of these approaches will have benefits and trade-offs, but all should be developed in partnership with city officials and community leaders to ensure public accountability, transparency, and ultimate benefit to the community.

Finally, data privacy and ethics should be considered in all stages of development and deployment. Sewage is a physical aggregate of multiple individuals, and so does not share many of the same re-identification concerns as human genomics and microbiome studies [29]. However, concerns about disparate impact and data misuse are real and should be addressed by bringing together community leaders, policy and privacy experts, and sociologists alongside the scientists and engineers developing these platforms for deployment.

Despite these challenges, sampling from upstream residential catchments has the potential to make wastewater testing more relevant to public health in cities. Residential catchments represent a geographic unit more amenable to public health programming and evaluation. By mitigating confounding factors across cities, they also have the potential to accelerate the discovery of population-level biomarkers and detection of geographical and temporal differences across urban communities. Within cities, residential sewage monitoring could alert public health officials of infectious disease outbreaks before a corresponding uptick in cases are seen in clinics and hospitals, allowing for earlier interventions in the affected communities. Mining residential sewage can also uncover behaviors not measurable by other means. For example, it could be possible to identify communities with high rates of non-lethal opioid overdoes by looking for locations with high naloxone or associated excretion metabolites but low reported overdose rates.

Because residential sewage may reflect a wide range of human behaviors, it could also be used for a wide range of policy applications. For example, analyzing the amounts of un-glucuronidated prescription opioids or antibiotics could be used to evaluate the effectiveness of drug take-back programs. Additionally, bacterial or chemical biomarkers of nutrition intake and diet could be used to study food deserts and the impact of programs increasing the availability of fresh foods in certain neighborhoods. Finally, marijuana metabolites and plant strains could be measured in sewage to measure the impact of cannabis legalization roll-outs across different counties or states.

6.4 Materials and Methods

6.4.1 Selection of Residential Catchment sampling site

The sampling site was identified by cross-referencing Cambridge's wastewater network maps, which contained the layout of pipes and manholes, as well as flow direction, with demographic and land use data. We selected a location where the upstream catchment fulfilled three major requirements: 1) a land use of over 80% residential so that we avoid transient populations and can characterize the demographics of the catchment population accurately, 2) a total catchment population >5,000 people to provide an anonymous reading of the community, and 3) the wastewater travel time from the furthest point in the catchment is <60 minutes to preserve the integrity of the sample.

6.4.2 Wastewater collection and processing

Grab wastewater samples (500 mL) were taken every hour from 10 AM on April 8th 2015 to 10 AM on April 9th 2015 (25 samples total). Samples were collected from the selected manhole with a commercial peristaltic pump (Global Water) sampling at a rate of 100 mL/min. 200 mL of collected sewage were filtered through 10-μm PTFE membrane filters. 150 ml of the outflow were filtered through 0.2-μm PTFE membrane filters. 100 ml of the final outflow were acidified with concentrated HCl (Optima grade, Fisher Scientific) to pH 3.0 and frozen at -80 degrees Celsius until metabolomics analysis. PTFE membrane filters were kept in RNAlater at -80 degrees Celsius until DNA extraction. The lab filtration system consisted of a Masterflex peristaltic pump (Pall), Masterflex PharMed BPT Tubing (Cole-Palmer), 47 mm PFA filter holders (Cole-Palmer) and 47mm PTFE Omnipore filter membranes (Millipore). Both the tubing and filter holders were previously cleaned with HCl (10% v/v) and ultrapure deionized water.

6.4.3 DNA extraction and amplicon based Illumina sequencing of 16S rRNA genes

0.2-μm filter membranes were thawed in ice. RNAlater was removed and filters were washed with phosphate-buffered saline (PBS) buffer twice. DNA was extracted from each filter with Power Water extraction kit (MO BIO Laboratories Inc.) according to manufacturer's instructions. The only modification was that DNA from up to three filters from the same sample was pooled into the same DNA-binding column. Paired-end Illumina sequencing libraries were constructed using a two-step PCR approach targeting 16S rRNA genes previously described [30]. All paired-end libraries were multiplexed into one lane and sequenced with paired end 300 bases on each end on the Illumina MiSeq platform at the MIT Biomicro Center.

6.4.4 Processing of 16S rRNA gene sequencing data

Raw data was processed with an in-house 16S processing pipeline (`https://github.com/thomasgurry/amplicon_sequencing_pipeline`). To assign OTUs, we clustered OTUs at 99% similarity using USEARCH [31] and assigned taxonomy to the resulting OTUs with the RDP classifier [32] and a confidence cutoff of 0.5. For each dataset, we removed samples with fewer than 100 reads and OTUs with fewer than 10 reads. We calculated the relative abundance of each OTU by dividing its value by the total reads per sample. We then collapsed OTUs by their RDP assignment up to genus level by summing their respective relative abundances.

6.4.5 Untargeted metabolomics analytical methods

Acetonitrile and deuterated biotin were added to the 0.2 μm-filtrate (acetonitrile final concentration was 5%, deuterated biotin was 0.05 μg ml^{-1}). The resulting solution was analyzed using liquid chromatography (LC) coupled via electrospray ionization (negative ion mode) to a linear ion trap-7T Fourier-transform ion cyclotron resonance hybrid mass spectrometer (Thermo Scientific, FT-ICR MS; LTQ FT Ultra). Chromatographic separation was performed using a Synergi Fusion C18 reversed phase

column (2.1 x 150mm, 4μm, Phenomenex). Samples were eluted at 0.25 mL/min with the following gradient: an initial hold at 95% A (0.1% formic acid in water) : 5% B (0.1% formic acid in acetonitrile) for 2 minutes, ramp to 65% B from 2 to 20 minutes, ramp to 100% B from 20 to 25 min, and hold at 100% B until 32.5 minutes. The column was re-equilibrated for 7 min between samples with solvent A. The injected sample volume was 20 μL. Full MS data were collected in the FT-ICR cell from m/z 100-1000 at 100,000 resolving power (defined at 400 m/z). In parallel to the FT acquisition, MS/MS scans of the four most intense ions were collected at nominal mass resolution in the ion trap (LTQ). Samples were analyzed in random order with a pooled sampled run every six samples in order to assess instrument variability.

6.4.6 Untargeted metabolomics raw data processing

Raw XCaliber files were converted to mzML files with MSConvert (`threshold` = 1000 for the negative ion mode files) (`http://proteowizard.sourceforge.net/tools.shtml`). Peaks were picked using the function `xcmsSet` from the R package `xcms`, with the following parameters (negative mode): `method` = 'centWave', `ppm` = 2, `snthresh` = 10, `prefilter` = (5, 1000), `mzCenterFun` = 'wMean', `integrate` = 2, `peakwidth` = (20, 60), `noise` = 1000, and `mzdiff` = -0.005 [33]. To align retention times, we used the `retcor.obiwarp` function with the following parameters: `plottype` = 'deviation', `profStep` = 0.1, `distFunc` = 'cor', `gapInit` = 0.3, and `gapExtend` = 0.4. To group peaks from different samples we used the `group.density` function with the following parameters: `minfrac` = 0, `minsamp` = 1, `bw` = 30, and `mzwid` = 0.001. To integrate areas of missing peaks, we used the `fillPeaks` function with `method` = 'chrom'.

6.4.7 Mass matching

The first step in identifying metabolites was comparing the exact mass values with masses of metabolites present in METLIN [34]. We searched METLIN for 397 of the most abundant features from the data and putatively identified 126 of these

159

based on m/z matches (ppm error $\leq$ 2). We also programmatically mapped all metabolite m/z values to the Human Metabolome Database (HMDB) database [35], https://github.com/cduvallet/match_hmdb). Untargeted m/z values were first converted to their expected neutral mass, assuming H^+ or Cl^- ions, and then scanned against the neutral masses of all HMDB compounds, with an error tolerance of 1 ppm. All HMDB compounds within the error tolerance were returned. When an m/z had multiple HMDB hits, the feature was named with the common chemical taxonomic classification. The HMDB database was downloaded in May 2016.

6.4.8   Glucuronides annotation through MS2 match

We attempted to confirm all putative glucuronides (from m/z hits through METLIN and HMDB searching) via matching of MS/MS data. MS/MS spectra were extracted from individual mzML files (via MZMine [36]) and matched to predicted in silico fragmentation patterns from MetFrag [37]. We considered a glucuronide to be confirmed if the expected glucuronide was among the top MetFrag hits and the observed fragments included peaks corresponding to the un-glucuronidated compound, diagnostic glucuronide derivatives (m/z = 113, 157, 175), or other diagnostic peaks (parent compound minus a CO_2 and/or OH).

### 6.4.9   Metabolite standards confirmation

We were able to confirm the putative identifications of several of the metabolites in Table 2 ('standard') at the most confident level using authentic standards analyzed with the LC-MS method described here. Each standard was analyzed individually in pure solvent as well as spiked into the 1 pm and 2 am samples. The expected m/z and retention time for each standard was determined from the standard-only runs and confirmed by identifying the corresponding m/z and retention time in the spike-in samples.

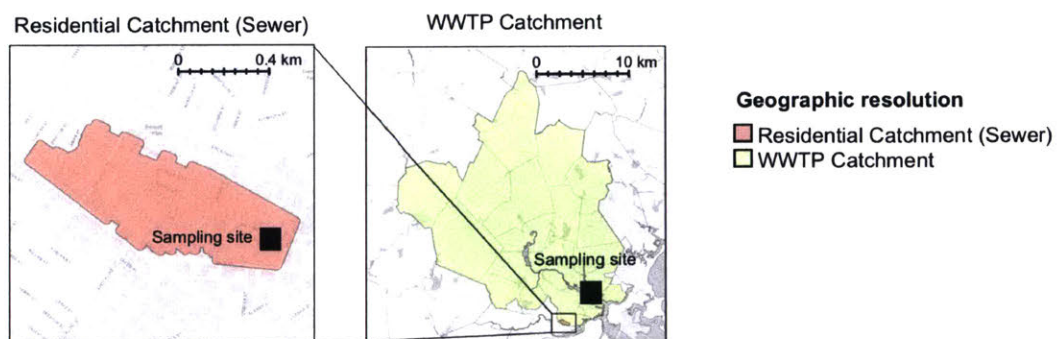## 6.5   Acknowledgements

6.6   Figures
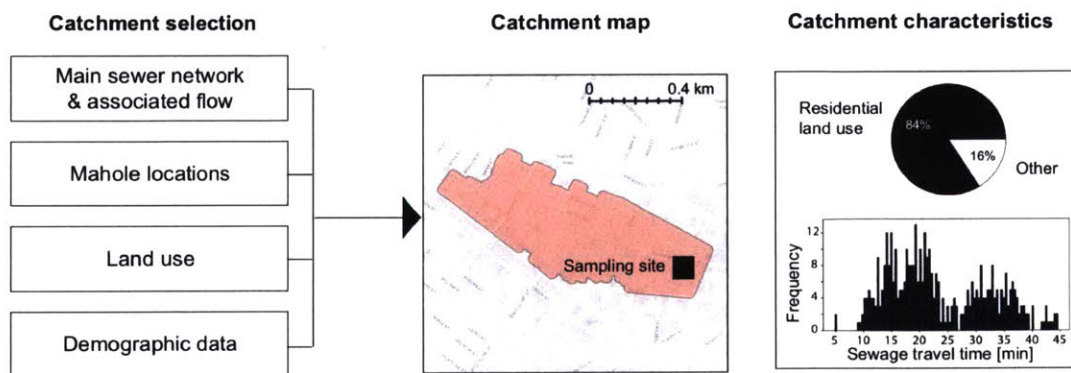
Figure 6-1: Definition and identification of residential catchments. (A). Integrating data from sewage network and flow, manhole locations, land usage, and demographic data allows for the selction of carefully curated residential manholes. (B) Upstream residential catchment vs. downstream catchment.
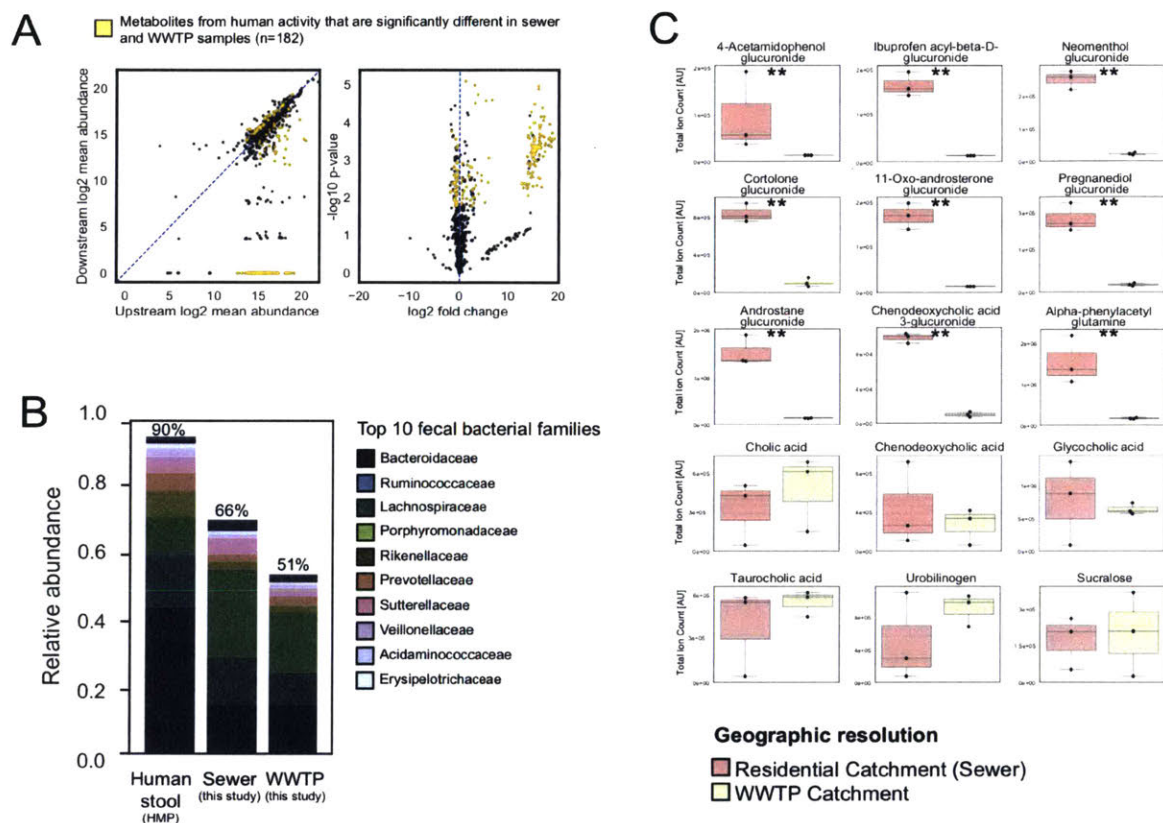
Figure 6-2: Upstream samples find more human-derived metabolites and bacteria. (A) Abundance of metabolite features in upstream vs. downstream sample (left) and volcano plot showing fold change vs. p-value (right). (B) Residential sewage sample ("Sewer") contains more human-associated bacteria than downstream sample ("WWTP"). (C) Abundance of various confirmed human-associated metabolites in upstream vs. downstream sample.
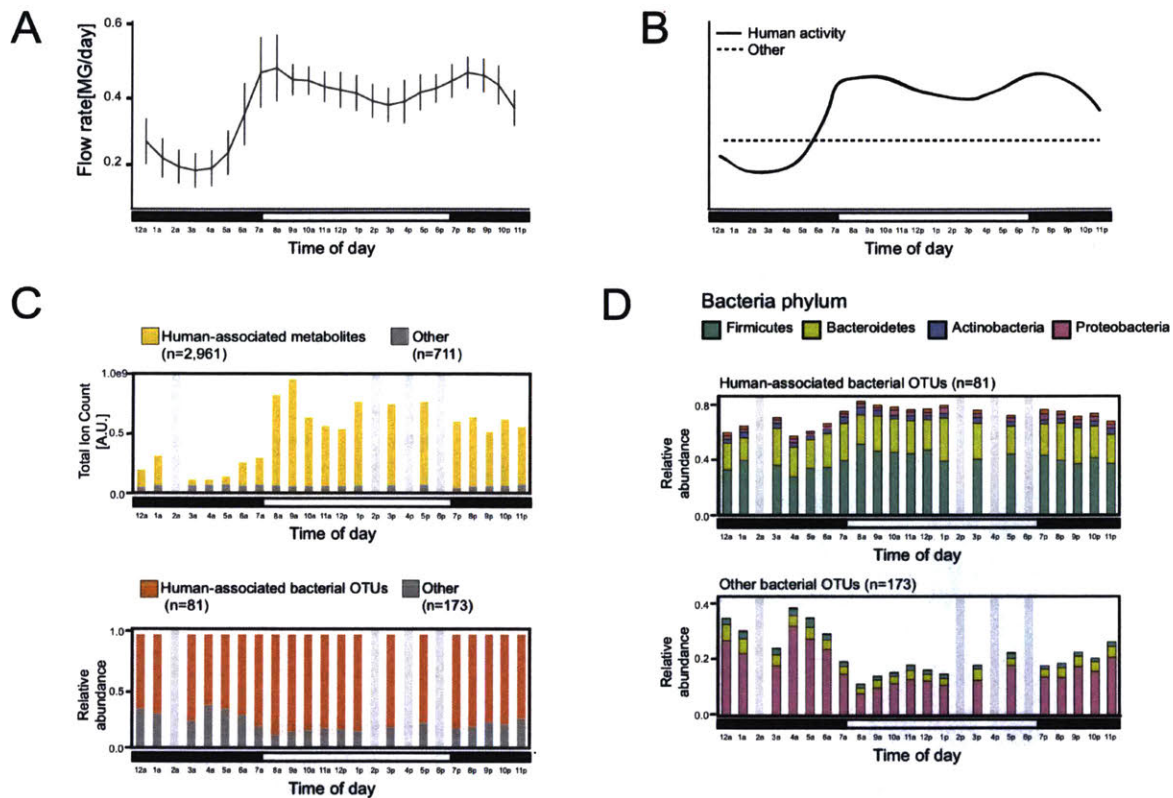
Figure 6-3: Human and environmental compartments reflect human diurnal dynamics. (A) Sampling campaign and schematic of expected human activity. (B) *Top*: average flow rate in residential manhole. *Bottom*: total ion count of M1 and M2 metabolite clusters. (C) Number of human-associated metabolites confirmed in M1 vs. M2 clusters. (D) *From top*: Average flow rate in residential manhole; relative abundance of two OTU clusters, O1 and O2; phylum-level taxonomic composition of bacterial cluster O1; phylum-level taxonomic composition of bacterial cluster O2.
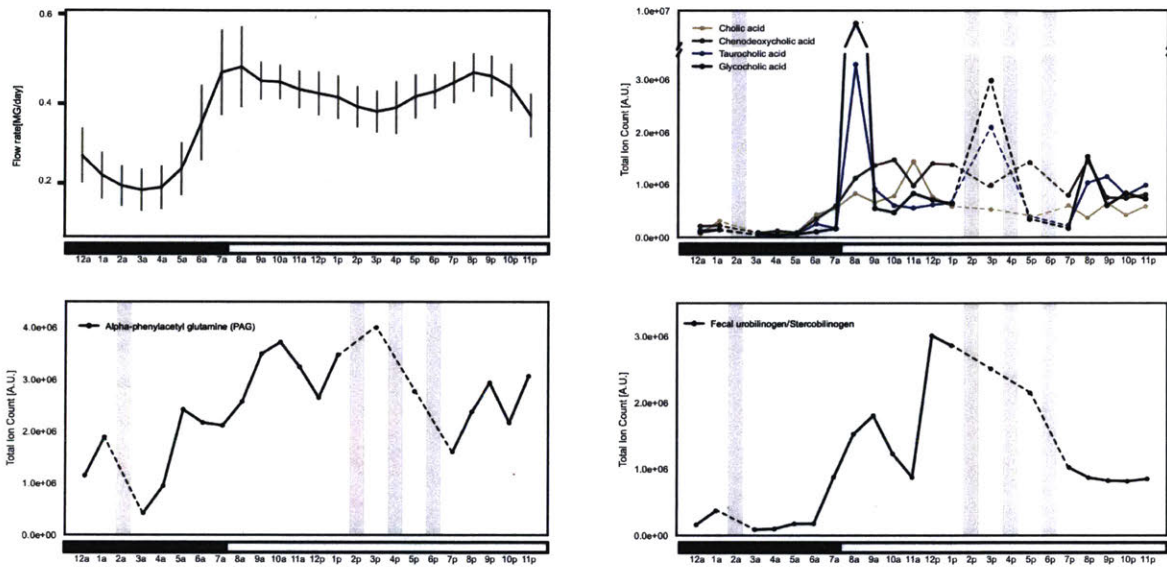
Figure 6-4: Confirmed urinary and fecal biomarkers reflect human activity and behavior. (Top left) Average flow rate in residential manhole. (Bottom left) Total ion count of alpha-phenylacetylglutamine (PAG), a confirmed human urinary marker. (Top right) Total ion count of confirmed bile acids, human fecal markers. (Bottom right) Total ion count of urobilinogen, a confirmed human fecal marker.

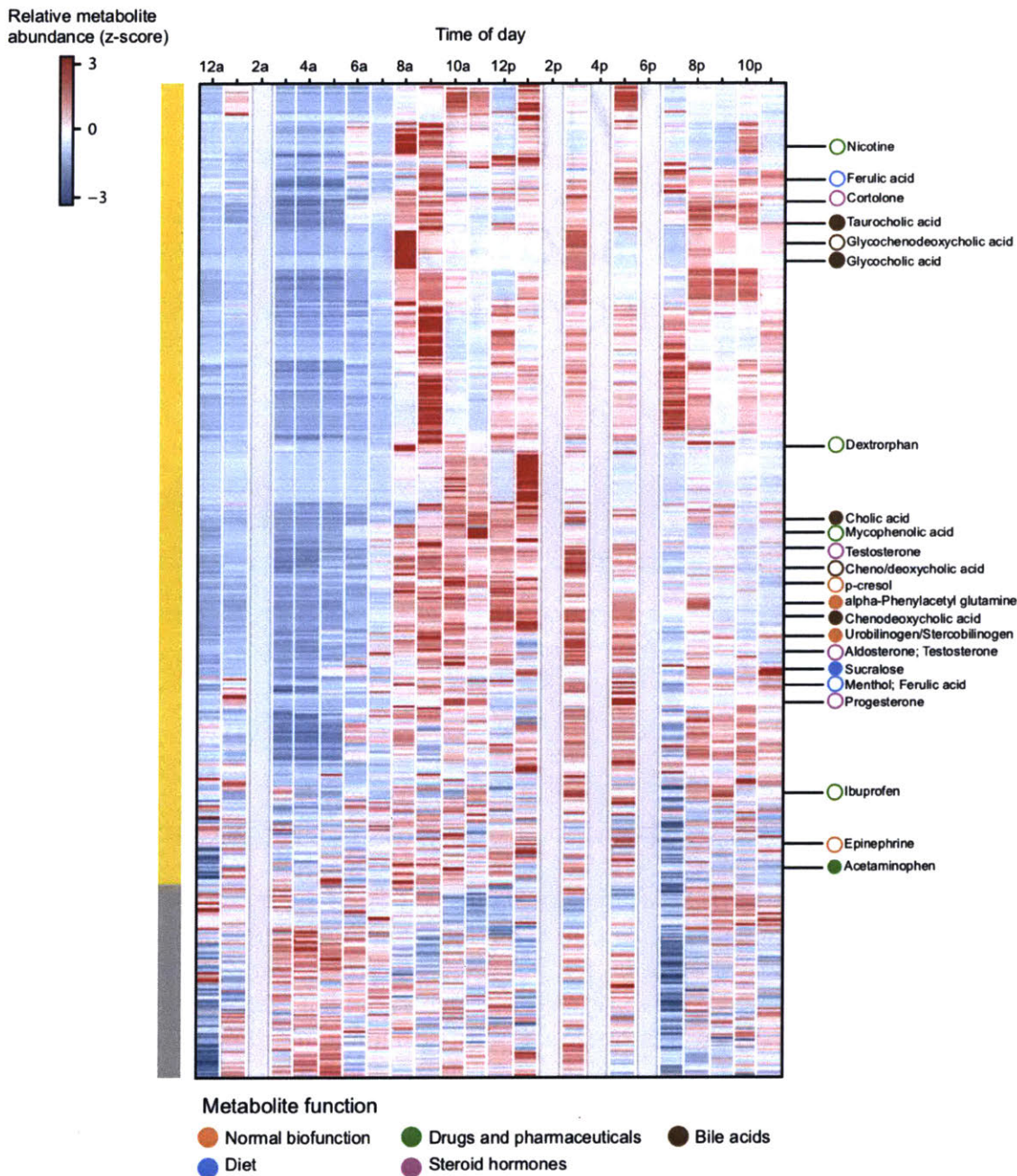Figure 6-5: Mining untargeted data suggests identification of human health, behavioral, and lifestyle biomarkers. Heatmap of metabolite feature abundance over 24-hour samplings, with putatively identified metabolites labeled. Metabolites labeled with filled circles were confirmed with analytical standards. Labels with open circles were identified via mass-matching neutral mass with HMDB metabolites and MS2 matching.

Bibliography

[1] Christian G. Daughton. Monitoring wastewater for assessing community health: Sewage chemical-information mining (SCIM). *Science of The Total Environment*, 619-620:748–764, apr 2018. doi: 10.1016/j.scitotenv.2017.11.102. URL https://doi.org/10.1016/j.scitotenv.2017.11.102.

[2] Y Manor, L M Shulma, E Kaliner, M Hindiyeh, D Ram, D Sofer, J Moran-Gilad, B Lev, I Grotto, R Gamzu, and E Mendelson. Intensified environmental surveillance supporting the response to wild poliovirus type 1 silent circulation in israel, 2013. *Eurosurveillance*, 19, 2014.

[3] Maria Hellmér, Nicklas Paxéus, Lars Magnius, Lucica Enache, Birgitta Arnholm, Annette Johansson, Tomas Bergström, and Heléne Norder. Detection of pathogenic viruses in sewage provided early warnings of hepatitis a virus and norovirus outbreaks. *Applied and Environmental Microbiology*, 80(21):6771–6781, aug 2014. doi: 10.1128/aem.01981-14. URL https://doi.org/10.1128/aem.01981-14.

[4] Ehud Kaliner, Eran Kopel, Emilia Anis, Ella Mendelson, Jacob Moran-Gilad, Lester M Shulman, Shepherd R Singer, Yossi Manor, Eli Somekh, Shmuel Rishpon, Alex Leventhal, Lisa Rubin, Diana Tasher, Mira Honovich, Larisa Moerman, Tamy Shohat, Ravit Bassal, Danit Sofer, Michael Gdalevich, Boaz Lev, Ronni Gamzu, and Itamar Grotto. The israeli public health response to wild poliovirus importation. *The Lancet Infectious Diseases*, 15(10):1236–1242, oct 2015. doi: 10.1016/s1473-3099(15)00064-x. URL https://doi.org/10.1016/s1473-3099(15)00064-x.

[5] Yakir Berchenko, Yossi Manor, Laurence S. Freedman, Ehud Kaliner, Itamar Grotto, Ella Mendelson, and Amit Huppert. Estimation of polio infection prevalence from environmental surveillance data. *Science Translational Medicine*, 9(383):eaaf6786, mar 2017. doi: 10.1126/scitranslmed.aaf6786. URL https://doi.org/10.1126/scitranslmed.aaf6786.

[6] Erica C. Pehrsson, Pablo Tsukayama, Sanket Patel, Melissa Mejía-Bautista, Giordano Sosa-Soto, Karla M. Navarrete, Maritza Calderon, Lilia Cabrera, William Hoyos-Arango, M. Teresita Bertoli, Douglas E. Berg, Robert H. Gilman, and Gautam Dantas. Interconnected microbiomes and resistomes

in low-income human habitats. *Nature*, 533(7602):212–216, may 2016. doi: 10.1038/nature17672. URL `https://doi.org/10.1038/nature17672`.

[7] Initiatives for addressing antimicrobial resistance in the environment: Current situation and challenges, 2018. URL `https://wellcome.ac.uk/sites/default/files/antimicrobial-resistance-environment-report.pdf`.

[8] Australian Criminal Intelligence Commission. National wastewater drug monitoring program – report 5, august 2018. 2018. URL `https://www.acic.gov.au/publications/intelligence-products/national-wastewater-drug-monitoring-program-report`.

[9] Alexander L.N. van Nuijs, Foon Yin Lai, Frederic Been, Maria Jesus Andres-Costa, Leon Barron, Jose Antonio Baz-Lomba, Jean-Daniel Berset, Lisa Benaglia, Lubertus Bijlsma, Dan Burgard, Sara Castiglioni, Christophoros Christophoridis, Adrian Covaci, Pim de Voogt, Erik Emke, Despo Fatta-Kassinos, Jerker Fick, Felix Hernandez, Cobus Gerber, Iria González-Mariño, Roman Grabic, Teemu Gunnar, Kurunthachalam Kannan, Sara Karolak, Barbara Kasprzyk-Hordern, Zenon Kokot, Ivona Krizman-Matasic, Angela Li, Xiqing Li, Arndís S.C. Löve, Miren Lopez de Alda, Ann-Kathrin McCall, Markus R. Meyer, Herbert Oberacher, Jake O'Brien, Jose Benito Quintana, Malcolm Reid, Serge Schneider, Susana Sadler Simoes, Nikolaos S. Thomaidis, Kevin Thomas, Viviane Yargeau, and Christoph Ort. Multi-year inter-laboratory exercises for the analysis of illicit drugs and metabolites in wastewater: Development of a quality control system. *TrAC Trends in Analytical Chemistry*, 103:34–43, jun 2018. doi: 10.1016/j.trac.2018.03.009. URL `https://doi.org/10.1016/j.trac.2018.03.009`.

[10] Kevin V. Thomas, Lubertus Bijlsma, Sara Castiglioni, Adrian Covaci, Erik Emke, Roman Grabic, Félix Hernández, Sara Karolak, Barbara Kasprzyk-Hordern, Richard H. Lindberg, Miren Lopez de Alda, Axel Meierjohann, Christoph Ort, Yolanda Pico, José B. Quintana, Malcolm Reid, Jörg Rieckermann, Senka Terzic, Alexander L.N. van Nuijs, and Pim de Voogt. Comparing illicit drug use in 19 european cities through sewage analysis. *Science of The Total Environment*, 432:432–439, aug 2012. doi: 10.1016/j.scitotenv.2012.06.069. URL `https://doi.org/10.1016/j.scitotenv.2012.06.069`.

[11] Jose Antonio Baz-Lomba, Stefania Salvatore, Emma Gracia-Lor, Richard Bade, Sara Castiglioni, Erika Castrignanò, Ana Causanilles, Felix Hernandez, Barbara Kasprzyk-Hordern, Juliet Kinyua, Ann-Kathrin McCall, Alexander van Nuijs, Christoph Ort, Benedek G. Plósz, Pedram Ramin, Malcolm Reid, Nikolaos I. Rousis, Yeonsuk Ryu, Pim de Voogt, Jorgen Bramness, and Kevin Thomas. Comparison of pharmaceutical, illicit drug, alcohol, nicotine and caffeine levels in wastewater with sale, seizure and consumption data for 8 european cities. *BMC Public Health*, 16(1), oct 2016. doi: 10.1186/s12889-016-3686-5. URL `https://doi.org/10.1186/s12889-016-3686-5`.

[12] Celia H. Arnaud. To monitor the health of citiesâĂŹ residents, look no further than their sewers. *Chemical & Engineering News*, 96, 2018.

[13] Ryan J. Newton, Sandra L. McLellan, Deborah K. Dila, Joseph H. Vineis, Hilary G. Morrison, A. Murat Eren, and Mitchell L. Sogin. Sewage reflects the microbiomes of human populations. *mBio*, 6(2), feb 2015. doi: 10.1128/mbio.02574-14. URL `https://doi.org/10.1128/mbio.02574-14`.

[14] A. Keshaviah, R. Gitlin, L. Cattell, W. Reeves, J. de Vallance, and C Thornton. The potential of wastewater testing for rapid assessment of opioid abuse (research brief). *Princeton, NJ: Mathematica Policy Research*, 2016. URL `https://www.mathematica-mpr.com/our-publications-and-findings/publications/the-potential-of-wastewater-testing-for-rapid-assessment-of-opioid-abuse-resea`

[15] A Keshaviah. The potential of wastewater testing for public health and safety. *Washington, DC: Mathematica Policy Research*, 2017. URL `https://www.mathematica-mpr.com/our-publications-and-findings/publications/the-potential-of-wastewater-testing-for-public-health-and-safety-special-report`

[16] Jake W. O'Brien, Phong K. Thai, Geoff Eaglesham, Christoph Ort, Andreas Scheidegger, Steve Carter, Foon Yin Lai, and Jochen F. Mueller. A model to estimate the population contributing to the wastewater using samples collected on census day. *Environmental Science & Technology*, 48(1):517–525, dec 2013. doi: 10.1021/es403251g. URL `https://doi.org/10.1021/es403251g`.

[17] Phong K. Thai, Guangming Jiang, Wolfgang Gernjak, Zhiguo Yuan, Foon Yin Lai, and Jochen F. Mueller. Effects of sewer conditions on the degradation of selected illicit drug residues in wastewater. *Water Research*, 48:538–547, jan 2014. doi: 10.1016/j.watres.2013.10.019. URL `https://doi.org/10.1016/j.watres.2013.10.019`.

[18] Christoph Ort, Michael G. Lawrence, JoĬĹrg Rieckermann, and Adriano Joss. Sampling for pharmaceuticals and personal care products (PPCPs) and illicit drugs in wastewater systems: Are your conclusions valid? a critical review. *Environmental Science & Technology*, 44(16):6024–6035, aug 2010. doi: 10.1021/es100779n. URL `https://doi.org/10.1021/es100779n`.

[19] Ettore Zuccato, Chiara Chiabrando, Sara Castiglioni, Renzo Bagnati, and Roberto Fanelli. Estimating community drug abuse by wastewater analysis. *Environmental Health Perspectives*, 116(8):1027–1032, aug 2008. doi: 10.1289/ehp.11022. URL `https://doi.org/10.1289/ehp.11022`.

[20] Sara Castiglioni, Ettore Zuccato, Elisabetta Crisci, Chiara Chiabrando, Roberto Fanelli, and Renzo Bagnati. Identification and measurement of illicit drugs and their metabolites in urban wastewater by liquid chromatography-tandem mass spectrometry. *Analytical Chemistry*, 78(24):8421–8429, dec 2006. doi: 10.1021/ac061095b. URL `https://doi.org/10.1021/ac061095b`.

[21] Rebecca M. Pollet, Emma H. D'Agostino, William G. Walton, Yongmei Xu, Michael S. Little, Kristen A. Biernat, Samuel J. Pellock, Loraine M. Patterson, Benjamin C. Creekmore, Hanna N. Isenberg, Rohini R. Bahethi, Aadra P. Bhatt, Jian Liu, Raad Z. Gharaibeh, and Matthew R. Redinbo. An atlas of β-glucuronidases in the human intestinal microbiome. *Structure*, 25(7):967–977.e5, jul 2017. doi: 10.1016/j.str.2017.05.003. URL https://doi.org/10.1016/j.str.2017.05.003.

[22] Alethea Jacox, Jillian Wetzel, Shu-Yuan Cheng, and Marta Concheiro. Quantitative analysis of opioids and cannabinoids in wastewater samples. *Forensic Sciences Research*, 2(1):18–25, jan 2017. doi: 10.1080/20961790.2016.1270812. URL https://doi.org/10.1080/20961790.2016.1270812.

[23] David S. Wishart, Timothy Jewison, An Chi Guo, Michael Wilson, Craig Knox, Yifeng Liu, Yannick Djoumbou, Rupasri Mandal, Farid Aziat, Edison Dong, Souhaila Bouatra, Igor Sinelnikov, David Arndt, Jianguo Xia, Philip Liu, Faizath Yallou, Trent Bjorndahl, Rolando Perez-Pineiro, Roman Eisner, Felicity Allen, Vanessa Neveu, Russ Greiner, and Augustin Scalbert. HMDB 3.0—the human metabolome database in 2013. *Nucleic Acids Research*, 41(D1):D801–D807, nov 2012. doi: 10.1093/nar/gks1065. URL https://doi.org/10.1093/nar/gks1065.

[24] METLIN: a metabolite mass spectral database. *Therapeutic drug monitoring*, 27:747–51, 2005.

[25] J.W.T. Seakins. The determination of urinary phenylacetylglutamine as phenylacetic acid. *Clinica Chimica Acta*, 35(1):121–131, nov 1971. doi: 10.1016/0009-8981(71)90302-0. URL https://doi.org/10.1016/0009-8981(71)90302-0.

[26] William H. Stein, Alejandro C. Paladini, C. H. W. Hirs, and Stanford Moore. PHENYLACETYLGLUTAMINE AS a CONSTITUENT OF NORMAL HUMAN URINE. *Journal of the American Chemical Society*, 76(10):2848–2849, may 1954. doi: 10.1021/ja01639a085. URL https://doi.org/10.1021/ja01639a085.

[27] B. Balikov. Urobilinogen excretion in normal adults: Results of assays with notes on methodology. *Clinical chemistry*, 3:145–153, 1957.

[28] Alan F. Hofmann. Enterohepatic circulation of bile acids. *Handbook of Physiology. The Gastrointestinal System.*, 4:567–596, 1989. doi: 10.1002/cphy.cp060329. URL https://doi.org/10.1002/cphy.cp060329.

[29] Eric A. Franzosa, Katherine Huang, James F. Meadow, Dirk Gevers, Katherine P. Lemon, Brendan J. M. Bohannan, and Curtis Huttenhower. Identifying personal microbiomes using metagenomic codes. *Proceedings of the National Academy of Sciences*, 112(22):E2930–E2938, may 2015. doi: 10.1073/pnas.1423854112. URL https://doi.org/10.1073/pnas.1423854112.

[30] Sarah P. Preheim, Allison R. Perrotta, Antonio M. Martin-Platero, Anika Gupta, and Eric J. Alm. Distribution-based clustering: Using ecology to refine the operational taxonomic unit. *Applied and Environmental Microbiology*, 79(21): 6593–6603, aug 2013. doi: 10.1128/aem.00342-13. URL https://doi.org/10.1128/aem.00342-13.

[31] R. C. Edgar. Search and clustering orders of magnitude faster than blast. *Bioinformatics*, 26(19):2460–2461, 2010. doi: 10.1093/bioinformatics/btq461. URL http://dx.doi.org/10.1093/bioinformatics/btq461.

[32] Qiong Wang, George M Garrity, James M Tiedje, and James R Cole. Naive bayesian classifier for rapid assignment of rrna sequences into the new bacterial taxonomy. *Applied and environmental microbiology*, 73(16):5261–5267, 2007.

[33] Colin A. Smith, Elizabeth J. Want, Grace O'Maille, Ruben Abagyan, and Gary Siuzdak. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, 78(3):779–787, feb 2006. doi: 10.1021/ac051437y. URL https://doi.org/10.1021/ac051437y.

[34] Ralf Tautenhahn, Kevin Cho, Winnie Uritboonthai, Zhengjiang Zhu, Gary J Patti, and Gary Siuzdak. An accelerated workflow for untargeted metabolomics using the METLIN database. *Nature Biotechnology*, 30(9):826–828, sep 2012. doi: 10.1038/nbt.2348. URL https://doi.org/10.1038/nbt.2348.

[35] David S. Wishart, Timothy Jewison, An Chi Guo, Michael Wilson, Craig Knox, Yifeng Liu, Yannick Djoumbou, Rupasri Mandal, Farid Aziat, Edison Dong, Souhaila Bouatra, Igor Sinelnikov, David Arndt, Jianguo Xia, Philip Liu, Faizath Yallou, Trent Bjorndahl, Rolando Perez-Pineiro, Roman Eisner, Felicity Allen, Vanessa Neveu, Russ Greiner, and Augustin Scalbert. HMDB 3.0 – the human metabolome database in 2013. *Nucleic Acids Research*, 41(D1):D801–D807, 2013. doi: 10.1093/nar/gks1065. URL https://doi.org/10.1093/nar/gks1065.

[36] Tomáš Pluskal, Sandra Castillo, Alejandro Villar-Briones, and Matej Orešič. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, 11(1): 395, 2010. doi: 10.1186/1471-2105-11-395. URL https://doi.org/10.1186/1471-2105-11-395.

[37] Sebastian Wolf, Stephan Schmidt, Matthias Müller-Hannemann, and Steffen Neumann. In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics*, 11(1):148, 2010. doi: 10.1186/1471-2105-11-148. URL https://doi.org/10.1186/1471-2105-11-148.

Chapter 7

Conclusions

7.1   Limitations and extensions of reported work

In this thesis, I present multiple projects which overcome a variety of challenges to mine human microbiome data to extract clinical insights. In the first project, I overcome the difficulty of identifying consistent biomarkers across highly variable lung and stomach communities by instead looking at the relationships between aerodigestive body sites within individual patients. In the second project, I re-analyze gut microbiome datasets across many diseases and individual studies to identify consistent associations despite the technical challenges involved in extracting generalizable knowledge from the existing corpus of published studies. This work motivated our development of a method to correct for batch effects presented in the third project. In the fourth project, I present a framework to categorize disease-specific insights gleaned from previously performed analyses and experiments and leverage them to improve fecal microbiota transplant clinical trials. Finally, in the fifth project I present preliminary results proposing residential sewage as a source of valuable public health information. Although performed in different contexts and with different goals in mind, these computational projects share similar limitations.

7.1.1   Associations not causation

First, the findings in each project suggest associations but do not and cannot confirm causal relationships. In Chapter 2, although I demonstrate that the interpretation of the aspiration-associated perturbation is consistent across a variety of analyses and metrics, such associations would still need to be recapitulated and confirmed in an independent patient cohort before being further developed as a clinical diagnostic. Also, many more mechanistic experiments and longitudinal cohort studies would need to be done to confirm their association with subsequent aspiration-related respiratory infections. In Chapter 3, the consistent associations that I identify across many different gut microbiome studies and microbiome-related diseases can make no claim of causality: in all case-control microbiome datasets, it is not possible to determine whether the microbiome shifts are a cause of the disease or simply responding to the host's physiological state. In fact, finding many non-specific disease- and health-associated bacteria across these datasets suggests that a large part of these associations are more likely to be responses to the host's general health status rather than specifically causal to any given disease. To truly find disease-associated bacteria, researchers need to identify consistent associations across multiple cohorts of their disease of interest and ensure that these associations are specific to that disease. To develop these into disease-specific therapeutics or confirm causal associations, researchers should go further and isolate the strains of interest to test their causality in animal models. Finally, the conceptual framework that I present in Chapter 5 suggests one way to advance translational microbiome science but will need to be applied in many FMT trials before its impact on FMT trial successes can be confirmed.

7.1.2   Data resolution limits insights and applicability

For the most part, these projects are anchored in the analysis of 16S rRNA data, which is more accessible than many 'omics data types but comes with its own set of inherent limitations. 16S data provides a window into which bacteria are present in microbial communities but does not indicate what functions these bacteria are performing. Be-

cause disease-mediating effects in individuals will result from bacterial function, future work should strive to incorporate function into their studies, for example by analyzing metagenomics, metabolomics or transcriptomics data, or a combination of these data types. I hope that as datasets become more readily available, a similar cross-disease meta-analysis as Chapter 3's will be performed, but based on metagenomics data rather than genus-level taxonomy. I expect that a function-focused meta-analysis will find much more consistent and readily interpretable associations within studies of the same disease and across multiple diseases.

Additionally, pairing 16S analyses with functional data could ensure that the bacterial DNA detected reflect the actual bacterial communities *living* in that sample. For example, it is not known to what extent bacteria detected in human lung and stomach microbiomes are living and actively growing vs. simply a readout of DNA from dead cells. Determining which bacteria are active in a community will be important to interpret findings and clinical interpretations from analyses of the human lung and stomach microbiomes, like the one presented in Chapter 2.

Another limitation of 16S data is that in most cases it cannot resolve bacteria at the strain level. Given that different strains of the same species have dramatically different clinical presentations, this is a crucial limitation in extracting clinically relevant associations from 16S microbiome data. The meta-analysis presented in Chapter 3 was performed at the genus level so that we could compare taxa from studies which sequenced different regions of the 16S gene. Genus-level analyses have much more limited biological interpretations than others performed at the species or strain-level. Even with the batch correction method developed in Chapter 4, meta-analyses will continue to be limited by the comparability of bacterial features between studies. Additionally, strain-level resolution will be especially important in evaluating rational donor selection methods, where engraftment of specific donor strains may be an important factor mediating patient responses to FMT. Finally, extending the work presented in Chapter 6 for public health applications in infectious disease surveillance will also necessarily require strain-level resolution to be useful.

7.1.3 Small sample sizes limit discoveries

Finally, these studies are limited by their sample sizes. Although the aerodigestive cohort presented in Chapter 2 and the meta-analysis in Chapter 3 are the largest of their kind to date, the analyses that I could perform were still often limited by insufficient samples. In Chapter 2, I was unable to draw robust conclusions about the relationship between reflux and the aerodigestive microbiome because the number of patients with the respective samples and metadata was too small to sufficiently power my analysis. In fact, throughout this project I frequently found my analyses limited by the necessary confluence of samples and metadata. For example, I would have liked to analyze the combinatorial effects of proton pump inhibitors, aspiration, and reflux on the aerodigestive microbiome, but I simply did not have enough patients with the respective samples and metadata to perform any reasonably powered analyses.

The meta-analysis presented in Chapter 3 was also limited by the number of datasets and diseases. Originally, I had hoped to compare disease-associated microbiome shifts in an unsupervised manner to identify patterns of shifts common to similar types of diseases. For example, I wondered if I could find a consistent group of bacteria associated with inflammation across multiple inflammatory diseases. However, I had a surprisingly difficult time finding datasets with publicly available raw data and associated patient-level clinical metadata, even given the large corpus of relevant published papers. Thus, I did not acquire a large enough variety of diseases and datasets to enable this broader categorization of diseases and perform statistically meaningful analyses on these groups of datasets. Future work could attempt these analyses on combined raw data, using the percentile-normalization method from Chapter 4 to correct for batch effects. However, such analysis will still be constrained by the number of studies which sequenced the same region of the 16S gene, since percentile-normalized data can only be combined across features (i.e. bacterial taxa) which are common across all datasets.

One of the main conclusions of the framework presented in Chapter 5 is that the usual sample sizes in FMT studies will almost always limit the power of retrospective

analyses to identify key bacteria. While a tempting response to this finding might be to expand the number of FMT patients in future clinical trials, this solution will usually not be practical for multiple logistical and ethical reasons. Unfortunately, because clinical studies are performed with human samples and sometimes require careful ethical justifications and considerable patient recruitment efforts, this challenge of small sample sizes is likely to remain an important limitation for future studies and clinical trials throughout the microbiome field.

As a pilot study evaluating the potential for residential sewage as a platform for wastewater epidemiology, the work presented in Chapter 6 had a small sample size by design. We performed this study to determine the feasibility of mining residential sewage for public health biomarkers, and powered it accordingly. At a broad level, the dynamics of human-associated metabolites followed the expected diurnal pattern of human activity, but any other patterns within this larger dynamic would need to be confirmed with larger studies. We also saw significant variability between different grab samples from the upstream site, suggesting that future studies should either aggregate sampling over a longer period of time or take enough replicates to achieve statistical confidence in any measured differences.

7.2 Re-analyzing existing datasets and data availability

A core theme that has emerged from my thesis work is that re-analyzing existing microbiome datasets can add substantial value to our field. One of my personal conclusions from the meta-analysis is that new cross-sectional gut microbiome studies should be regularly contextualized within the existing published body of work, almost as mini meta-analyses of their own. Now that microbiome datasets are more readily available and bioinformatics tools to process and analyze them are becoming very accessible, I hope that researchers make it a habit to ask: are these associations replicated in independent patient cohorts? And: are they specific to my disease of

interest or are they part of a non-specific response to health and disease? I hope that our batch correction method can help researchers answer these questions, and that more statistical and bioinformatics tools will be developed to make meta-analyses more accessible and powerful for future researchers.

I also found the value-add of re-analyzing datasets especially relevant in microbiome research led by clinicians. For example, the original IBD FMT trials that I re-analyzed in Chapter 5 did not thoroughly investigate potential donor effects on patient response in their original studies. This is expected in part because donor heterogeneity was not usually a primary research question of interest, and also because the trials were not powered for such analyses. However, now that multiple IBD FMT trials have been published with paired patient and donor microbiome sequencing data, hypotheses about what leads to improved patient response can be tested *in silico* (as we did for butyrate producer abundance in Chapter 5). In Chapter 3, many of my datasets were pulled from studies led by clinical researchers who may not have had access to the bioinformatics and statistical expertise to fully interrogate disease associations within their datasets. By making their data publicly available, their work continued to contribute new knowledge to the field.

Moving forward, I hope and expect that researchers continue to make their raw data publicly available and that re-analyses of such data become standard in the field. Publicly available raw data allows researchers to ask and answer new questions, testing their hypotheses *in silico* without the need for costly new studies. New studies can also be compared with existing work to determine which of their findings hold across different studies and which are specific to their specific study. However, using published data in new contexts comes with challenges. As described in Chapter 4, batch effects resulting from different experimental and sequencing methods can make it very difficult to compare data across different labs. Another major challenge is data availability, and specifically clinical metadata like patient disease diagnosis. Through this work, I have come to realize that raw data without its associated metadata is in almost all cases useless. Given these challenges, standards for sharing data which respect patient privacy, clinicians' efforts for patient recruitment, and the needs of

computational biologists will need to be agreed upon and upheld as a community.

7.3 Partnerships between practitioners and computational biologists

Throughout this thesis, I also came to appreciate the unique contributions that come from close partnerships between clinicians and computational biologists. None of the projects in this thesis would have been possible without crucial contributions from our clinical counterparts. Chapter 2 was only possible because our clinical PI, Rachel Rosen, identified and framed the questions of interest in this cohort. Then, I was able to translate her questions into computational analyses and provide preliminary answers with our data. Working together in this way, we discovered new science and found clinically exciting results. The meta-analysis in Chapter 3 was also significantly strengthened by the inclusion of datasets which were originally purely clinical investigations, and the framework presented in Chapter 5 was only possible because of our lab's strong ties with the clinical experts at OpenBiome. I hope that the future of translational microbiome research establishes structures that encourage such close collaborations. Systems to share raw data should be designed with these collaborations in mind: the process to deposit data should be accessible to clinicians, patient information should be protected while also providing easy access to analyses that don't use the protected information, and the metadata should be curated well enough to enable straightforward analyses without much manual curation but also flexible enough to allow for the variety of study designs pursued by clinicians.

I was also especially impressed by the unique power of collaborations between scientists and practitioners through my involvement in the work presented in Chapter 6. That project was a result of coordination between multiple scientific disciplines as well as our city's public works and public health departments. The urban designers on our team incorporated our computational biology perspectives into a larger vision of the future of "smart cities." The public health officials we worked with helped us

understand and address their practical needs, and also facilitated discussions with the community to ensure that we were being transparent and locally engaged. Finally, our conversations with public works employees like Herbie as we stood by open manholes during our sampling campaigns gave us important insights to contextualize our experimental results and were a unique addition to my PhD experience.

7.4   Finding knowledge in information

A turning point in my thesis came when I read Gene Glass's 1976 paper coining the term "meta-analysis" [1]. In it, Glass argues for the underappreciated importance of consolidating and synthesizing information into knowledge, prizing work that aims to find meaning and draw conclusions from disparate existing studies. This was a turning point in my thesis for two reasons. First, it was the moment I became truly proud of my work, especially the meta-analysis in Chapter 3: I realized that it was not just some sort of microbiome "stamp collection" endeavor that anyone with basic knowledge of computational tools could do, but was instead difficult and valuable work that I had uniquely contributed to. After reading this piece, my perspective on my thesis work changed: before, I sometimes felt that the inherent limitations of computational work meant that projects like mine were not quite as valuable as theses which develop novel experimental methods or generate new data directly testing biological hypotheses. Now, I recognize that they are also invaluable work on their own merit. Second, reading Glass's words helped me recognize a uniting theme in all of my work: finding the "knowledge in the information." I realized that in each project presented in this thesis, I had not just mined the data to find statistical associations, but rather to interpret the associations I found and chase them all the way to their potential clinical or public health implications. I hope that as we move forward in this exciting and vibrant field, microbiome researchers collectively become less satisfied with simply reporting new information, instead emphasizing and valuing efforts that synthesize existing knowledge and generate new insights that lead more directly to clinical impact.

Bibliography

[1] Gene V. Glass. Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10):3–8, nov 1976.

Appendix A

Supplementary Information for Chapter 2

A.1 Supplementary Tables

Family	Genus	Non-aspirator	Aspirator	Difference
Neisseriaceae	Neisseria	7.1	41.4	34.2
Porphyromonadaceae	Porphyromonas	28.6	62.1	33.5
Pasteurellaceae	Haemophilus	50.0	82.8	32.8
Lachnospiraceae	Coprococcus	10.7	37.9	27.2
Micrococcaceae	Rothia	14.3	41.4	27.1
Prevotellaceae	Prevotella	25.0	51.7	26.7
Carnobacteriaceae	Granulicatella	32.1	58.6	26.5
Bacillales_Incertae_Sedis_XI	Gemella	42.9	69.0	26.1
Pasteurellaceae	Haemophilus	57.1	82.8	25.6
Actinomycetaceae	Actinomyces	17.9	41.4	23.5
Streptococcaceae	Streptococcus	39.3	62.1	22.8
Lachnospiraceae	Oribacterium	14.3	34.5	20.2
Leptotrichiaceae	Streptobacillus	17.9	37.9	20.1
Lachnospiraceae	Lachnoanaerobaculum	17.9	37.9	20.1
Fusobacteriaceae	Fusobacterium	42.9	62.1	19.2
Prevotellaceae		50.0	69.0	19.0
Flavobacteriaceae	Planobacterium	14.3	31.0	16.7
Leptotrichiaceae	Leptotrichia	14.3	31.0	16.7
Erysipelotrichaceae	Solobacterium	17.9	34.5	16.6
Prevotellaceae	Prevotella	21.4	37.9	16.5
Pasteurellaceae	Haemophilus	28.6	44.8	16.3
Veillonellaceae	Veillonella	35.7	51.7	16.0

Family	Genus	Non-aspirator	Aspirator	Difference
Prevotellaceae		46.4	62.1	15.6
Enterobacteriaceae	Escherichia/Shigella	46.4	62.1	15.6
Neisseriaceae	Neisseria	60.7	75.9	15.1
Streptococcaceae	Streptococcus	75.0	89.7	14.7
Veillonellaceae	Veillonella	35.7	48.3	12.6
Micrococcaceae	Rothia	42.9	55.2	12.3
Streptococcaceae	Streptococcus	42.9	55.2	12.3
Prevotellaceae	Prevotella	42.9	55.2	12.3
Prevotellaceae	Prevotella	64.3	75.9	11.6
Unknown Burkholderiales		10.7	20.7	10.0
Bacteroidaceae	Bacteroides	14.3	24.1	9.9
Porphyromonadaceae	Porphyromonas	21.4	31.0	9.6
Moraxellaceae	Moraxella	39.3	48.3	9.0
Prevotellaceae	Prevotella	57.1	65.5	8.4
Leptotrichiaceae	Leptotrichia	21.4	27.6	6.2
Fusobacteriaceae	Fusobacterium	25.0	31.0	6.0
Porphyromonadaceae	Porphyromonas	50.0	55.2	5.2
Neisseriaceae	Neisseria	17.9	20.7	2.8
Veillonellaceae	Veillonella	89.3	89.7	0.4
Unknown Bacteria		21.4	20.7	-0.7
Coriobacteriaceae	Atopobium	21.4	20.7	-0.7
Enterococcaceae		85.7	82.8	-3.0
Chloroplast	Streptophyta	10.7	6.9	-3.8
Pasteurellaceae	Haemophilus	17.9	13.8	-4.1
Unknown Bacillales		17.9	13.8	-4.1
Lactobacillaceae	Lactobacillus	28.6	17.2	-11.3
Pasteurellaceae	Haemophilus	32.1	20.7	-11.5
Staphylococcaceae	Staphylococcus	60.7	48.3	-12.4
Comamonadaceae	Acidovorax	17.9	3.4	-14.4
Porphyromonadaceae	Parabacteroides	21.4	6.9	-14.5
Comamonadaceae	Pelomonas	21.4	6.9	-14.5
Flavobacteriaceae	Chryseobacterium	28.6	13.8	-14.8
Erysipelotrichaceae	Clostridium_XVIII	21.4	3.4	-18.0
Lachnospiraceae	Ruminococcus2	25.0	6.9	-18.1
Flavobacteriaceae	Chryseobacterium	50.0	31.0	-19.0
Neisseriaceae	Microvirgula	57.1	37.9	-19.2
Enterobacteriaceae	Enterobacter	82.1	62.1	-20.1
Mycobacteriaceae	Mycobacterium	28.6	6.9	-21.7
Moraxellaceae	Acinetobacter	60.7	37.9	-22.8
Streptococcaceae	Streptococcus	60.7	37.9	-22.8
Bacteroidaceae	Bacteroides	53.6	27.6	-26.0
Unknown Bacillales		57.1	31.0	-26.1
Moraxellaceae	Acinetobacter	60.7	34.5	-26.2
Moraxellaceae	Enhydrobacter	60.7	34.5	-26.2

Family	Genus	Non-aspirator	Aspirator	Difference
Lactobacillaceae	Lactobacillus	50.0	20.7	-29.3
Aeromonadaceae	Aeromonas	57.1	27.6	-29.6
Moraxellaceae	Acinetobacter	78.6	41.4	-37.2
Leuconostocaceae	Leuconostoc	78.6	37.9	-40.6
Leuconostocaceae	Weissella	78.6	37.9	-40.6
Moraxellaceae	Acinetobacter	78.6	37.9	-40.6
Streptococcaceae	Lactococcus	78.6	37.9	-40.6
Streptococcaceae	Lactococcus	78.6	37.9	-40.6

Table A.1: Prevalence of lung-gastric fluid exchanged OTUs. Prevalence is calculated as the percentage of patients who have the OTU present in both their lungs and oropharynx, calculated separately among aspirators (N = 29) and non-aspirators (N = 28). OTUs are ordered by their differential prevalence in aspirators relative to non-aspirators, and are labeled with their family- and genus-level taxonomies. Blank genus names indicate OTUs which were not annotated at the genus level.

Lung-oropharynx OTUs (12)	AUC	p	N (non-asp/asp)
Lung	0.60	0.32	33/33
Oropharyngeal	0.64	0.13	43/36
Both	0.73	0.14	23/25
Lung-gastric OTUs (75)			
Lung	0.61	0.42	33/33
Gastric fluid	0.68	0.04	48/41
Both	0.71	0.07	28/29

Table A.2: **Classifiers based on the abundance of exchanged OTUs.** (Top) Classifiers built from the abundance of lung-oropharynx exchanged OTUs. (Bottom) Classifiers built from the abundance of lung-gastric exchanged OTUs. Rows indicate which microbial community was used to train each classifier. In classifiers using two sites ("Both"), abundances of each exchanged OTU in each site were considered as separate features. AUCs are calculated as the area under the average ROC curve from five-fold cross validation. Fisher's exact p values are calculated on the predictions on the hold-out data for all cross validation folds using Python's `scipy.stats.fisher_exact` function. Each classifier was built 100 times with random patient splits and classifier initializations, and mean values are reported here. AUCs and Fisher p-values from all 100 repetitions for all classifiers are shown in Supplementary Figures A-2 and A-3.)
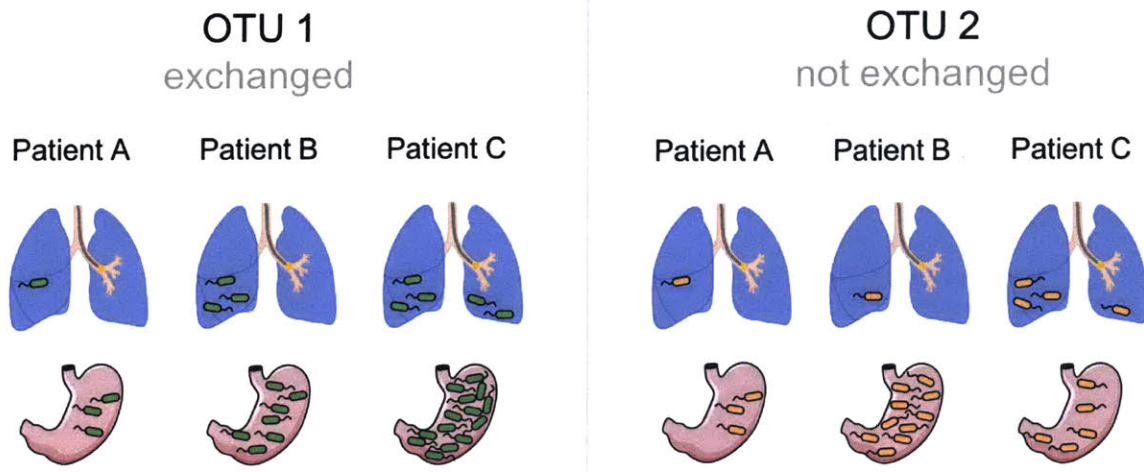
A.2 Supplementary Figures

Figure A-1: Schematic illustrating an OTU which is considered exchanged between the lung and stomach (left) and one which is not (right). If an OTU is exchanged in two sites, its abundance in the two sites should be correlated across patients. Lung image was adapted from Cancer Research UK / Wikimedia Commons and the stomach image from Servier Medical Art.
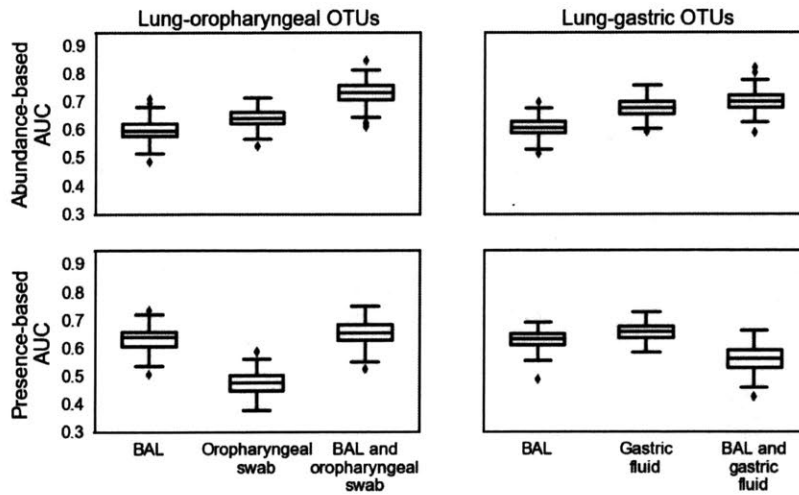
Figure A-2: Areas under the ROC curve (AUC) for 100 classifiers trained on the abundance (top) or presence (bottom) of lung-oropharynx exchanged OTUs (left) or lung-gastric exchanged OTUs (right).
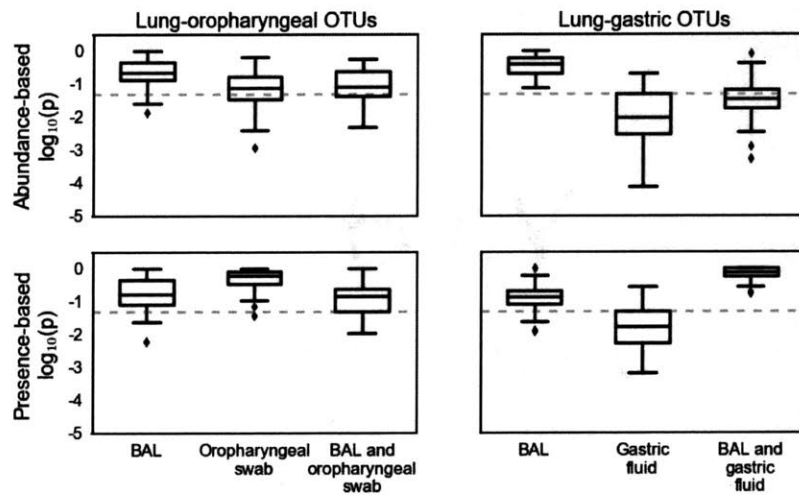

Figure A-3: Log of the Fisher p-values for 100 classifiers trained on the abundance (top) or presence (bottom) of lung-oropharynx exchanged OTUs (left) or lung-gastric exchanged OTUs (right). Dashed line indicates p = 0.05.
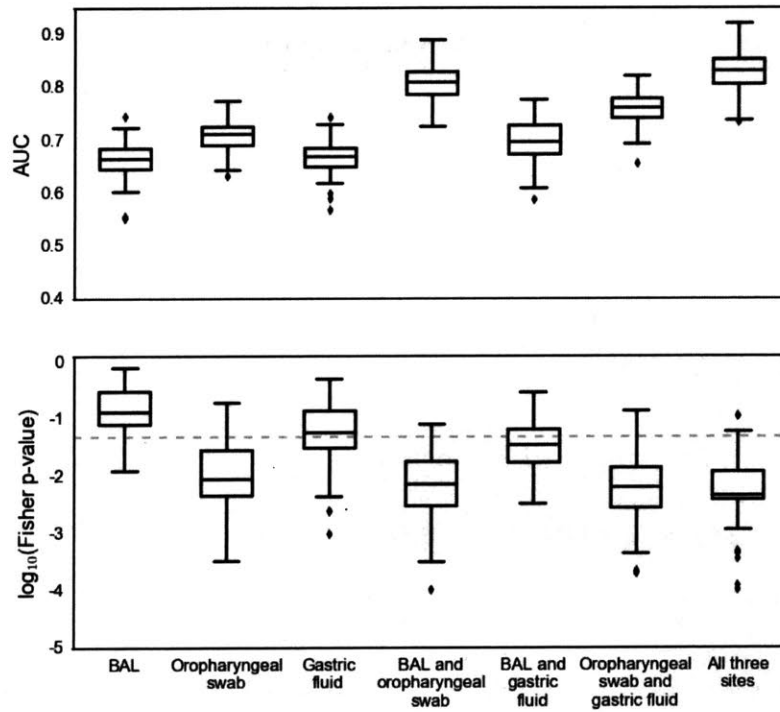
Figure A-4: Area under the ROC curve (AUC) (top) and Fisher p-values (bottom) for 100 classifiers trained on different combinations of the full aerodigestive communities to distinguish aspirators from non-aspirators. Dashed line on the p value plot is p = 0.05.

Appendix B

Supplementary Information for Chapter 3

B.1 Supplementary Notes

B.1.1 Re-processing and re-analyzing raw data yields results which are generally consistent with previously published results

Our re-analyses of the 29 studies were largely consistent with the originally reported results, with the same taxonomic groups showing similar trends despite differences in data-processing methodologies. We usually found fewer significant ($q < 0.05$) differences between control and diseased groups, which is likely due to our choice of a non-parametric statistical test (Kruskall-Wallis) paired with a multi-test correction (FDR). Thus, our results are more conservative. We also collapsed to genus level in order to compare results across disparate studies, which prevented us from identifying species- or strain-specific associations which the original authors may have identified. A major advantage of our re-analysis is that each data set was processed and analyzed in the same way, which allowed us to more directly compare results across studies and diseases.

Clostridium difficile **Infection and enteric diarrhea are characterized by large-scale shifts in the microbiome (CDI; 4 studies)**

Schubert et al. (2014) looked at how the gut microbiota differed between CDI patients with diarrhea (n = 94), non-CDI patients with diarrhea (n = 89), and non-diarrheal controls (n = 155) [1]. Similar to other CDI studies, the authors found a significant reduction in alpha diversity in patients with diarrhea (DunnâĂŹs multiple-comparison test on AMOVA, p < 0.0001). They found that OTUs from the *Ruminococcaceae, Lachnospiraceae, Bacteroides, Prevotellaceae,* and *Porphyromonadaceae* families were enriched in healthy subjects relative to patients with CDI and non-CDI diarrhea. They also showed that OTUs from the *Enterococcus* genus and the *Enterobacteriaceae* and *Erysipelotrichaceae* families were more prevalent in patients with diarrhea. In our analysis of the data, we also observed a significant reduction in alpha diversity in patients with diarrhea (q <= 0.05, KW test). Similarly, we found that *Enterobacteriaceae, Enterococcus,* and *Erysipelotrichaceae* were enriched in CDI patients, in addition to *Fusobacterium, Parvimonas, Veillonella, Carnobacterium, Streptococcus, Tetragenococcus, Lactobacillus, Pediococcus, Gemella, Staphylococcus, Butyricicoccus, Robinsoniella, Clostridium XlVa, Clostridium XlVb, Ruminococcus2, Flavonifractor, Gemmiger, Mogibacterium, Peptostreptococcus, Clostridium XI, Eggerthella, Atopobium, Actinomyces, Arthrobacter, Aggregatibacter, Pseudomonas,* and *Dysgonomonas.* As in the original study, we found that *Bacteroides, Alistipes, Anaerovorax, Oxalobacter, Bordetella, Prevotellaceae, Porphyromonadaceae, Lachnospiraceae,* and *Ruminococcaceae* were more abundant in the healthy controls. We also found *Turicibacter, Dialister, Eubacterium, Asteroleplasma, Cloacibacillus, Bordetella, Oxalobacter, Sutterella, Parasutterella, Desulfovibrio, Sediminibacterium,* and *Methanobrevibacter* to be enriched in the controls (q <= 0.05, KW tests). Overall, our analysis closely matched what was presented in the original manuscript.

Vincent et al. (2013) compared 25 patients with CDI to 25 healthy control patients [2]. The authors found a significant reduction in alpha diversity (p <= 0.05, Mann-Whitney U test). They also report a reduction in *Bacteroidaceae* and *Clostridiales*

189

Incertae Sedis XI in CDI patients relative to controls, and an enrichment in *Enterococcaceae* in CDI patients (p < 0.05, logistic regression). After reprocessing these data and collapsing abundances to the genus level, we observed a similar reduction in alpha diversity (q <= 0.05, KW test). We saw that the *Enterococcaceae* genera *Enterococcus* and *Proteus* were enriched in CDI patients. Healthy controls showed higher levels of *Fusobacterium, Peptoniphilus, Murdochiella, Anaerococcus, Finegoldia, Odoribacter, Prevotella,* and *Parabacteroides* relative to CDI patients. In summary, our results are fairly similar to the authors' original analysis, showing a depletion in *Bacteroidetes* and an enrichment in *Proteobacteria* in CDI patients.

Youngster et al. (2014) applied fecal microbiota transplants (FMTs) with materials collected from 5 healthy donors to 20 patients with recurrent *Clostridium difficile* infections (CDIs) [3]. The goal of this study was to determine whether nasal-gastric tube or colonoscopy administration of FMTs was most effective for treating CDIs (i.e. half of the CDI patients received one or the other treatment). The authors reported a significant reduction in alpha diversity in CDI patients vs. the healthy donors (p < 0.001, Mann-Whitney test). They did not assess whether there were significant differences in microbial community composition between CDI patients and donors, although they show that composition becomes more similar to donors following FMT. In our analysis, we also found a significant reduction in alpha diversity (p <= 0.05, KW test). *Enterococcus* was enriched in CDI patients relative to healthy stool donors (q <= 0.05, KW tests) and 15 genera were depleted in CDI patients relative to healthy controls. Healthy donors were enriched in genera from *Ruminococcaceae* and *Lachnospiraceae* families, in addition to the genera *Dialister* and *Anaerosporobacter*.

Singh et al. (2015) examined differences in the gut microbiome between individuals with enteric infections (n=200) and healthy controls (n=75) [4]. The authors report a significant drop in alpha diversity in diseased patients relative to the controls (unknown test). They also report a general reduction in the dominance of *Firmicutes* and *Bacteroidetes* phyla and an increase in the prevalence of *Proteobacteria* in diseased patients. Specifically, they report an increase in the abundance of *Enterobacteriaceae, Lactobacillaceae, Pasteurellaceae, Streptococcus, Bacilli, Escherichia,*

Haemophilus, and certain *Ruminococcus* species in patients with diarrhea. In healthy people, they report a significant enrichment in *Verrucomicrobia, Dorea, Blautia, Holdermania, Ruminococcaceae, Lachnospiraceae, Butyricimonas, Faecalibacterium, Bacteroidaceae*, and *Bifidobacterium, Sutterella, Parabacteroides, Rikenellaceae*, and *Oscillospira*. After re-processing the data, we found very similar results to those originally reported. We found that alpha diversity was significantly lower in patients with enteric infections (q <= 0.05, KW test). We saw significant enrichment in *Proteobacteria* families in patients with diarrhea, including *Enterobacteriaceae, Pasteurellaceae, Campylobacteraceae*, and *Neisseriaceae*. We also saw higher levels of *Fusobacterium, Parvimonas, Veillonella, Lactococcus, Streptococcus, Enterococcus, Tetragenococcus, Gemella, Ruminococcus II, Peptostreptococcus*, and *Collinsella* in diseased patients. In the healthy controls, we found enrichment of 43 genera, including *Sutterella, Verrucomicrobia (Akkermansia), Ruminococcaceae, Lachnospiraceae, Bacteroidaceae*, and *Bifidobacterium*. In addition, we saw higher levels of several members of *Rumminococcaceae, Lachnospiraceae*, and *Bacteroidales* in healthy controls (q <= 0.05, KW tests). Overall, our results largely overlap with those presented, but we identify a number of significant taxa that were not originally reported.

Taken together, we see large-scale shifts in the microbiome associated with both CDI and non-CDI diarrhea. The dysbiosis of enteric infection and diarrhea is quite consistent across studies. In general, *Proteobacteria* increase in prevalence in patients with diarrhea, with a concomitant decrease in *Bacteroidetes* and *Firmicutes*. In particular, we see a reduction in butyrate-producing Clostridia, including genera within *Ruminococcaceae* and *Lachnospiraceae* families, which have been associated with a healthy gut. We also see in increase in prevalence of organisms often associated with lower pH and higher oxygen levels of the upper-gut, like *Lactobacillaceae* and *Enterobacteriaceae* [5], in patients with diarrhea. Thus, diarrhea leads to consistent and large-scale rearrangements in the composition of the gut microbiome.

Colorectal Cancer has a consistent, ootentially pathogenic microbial signature (CRC; 4 studies)

Baxter et al. (2016) looked at differences in the microbiomes of 120 colorectal cancer (CRC) patients, 198 patients with non-cancerous adenomas, and 172 healthy controls [6]. Similar to prior work, the authors found that *Porphyromonas, Peptostreptococcus, Parvimonas*, and *Fusobacterium* were positively associated with CRC (random forest classifiers). Furthermore, they found that the absence of certain *Lachnospiraceae* species was associated with the presence of adenomas. We found similar patterns in our re-analysis of these data, with *Fusobacterium, Peptostreptococcus, Parvimonas*, and *Porphyromonas* enriched in CRC patients (q <= 0.05, KW tests). We also found higher levels of *Victivallis, Peptoniphilus, Anaerococcus, Catenibacterium, Staphylococcus, Collinsella, Enterobacter*, and *Alloprevotella* in CRC patients (q <= 0.05, KW tests). We found that healthy controls were enriched in *Lachnobacterium* (genus within *Lachnospiraceae*), *Gemmiger* (within *Rumminococcaceae*), *Clostridium XVIII*, and *Haemophilus* (q <= 0.05, KW tests). Overall, these results match what has been reported previously for CRC [7].

Zeller et al. (2014) collected microbiome data from 41 CRC patients and 75 control patients [8]. At the phylum level, they found that *Proteobacteria, Fusobacteria*, and *Bacteroidetes*, were more abundant in CRC patients, while *Firmicutes* and *Actinobacteria* were enriched in control patients. At the genus level, the authors report higher levels of *Fusobacterium, Pseudoflavonifractor, Peptostreptococcus, Leptotrichia, Porphyromonas, Desulfovibrio, Parvimonas, Selenomonas*, and *Bilophila* in CRC patients (q <= 0.1, FDR-corrected Wilcoxon tests). Healthy controls were enriched in *Bifidobacterium, Acinetobacter, Campylobacter, Ruminococcus*, and *Eubacterium* genera (q <= 0.1, FDR-corrected Wilcoxon tests). In our re-analysis we found enrichment of *Fusobacterium, Parvimonas, Flavonifractor, Anaerotruncus, Anaerovorax, Peptostreptococcus, Comamonas, Eikenella, Butyricimonas*, and *Porphyromonas* genera in CRC patients (q <= 0.05, KW tests). In healthy patients, we found higher levels of *Anaerostipes* (within *Lachnospiraceae*; q <= 0.05, KW tests).

Wang et al. (2011) analyzed a cohort of 46 CRC patients and 56 healthy controls [9]. The authors found no difference in alpha diversity between CRC and control patients. CRC patients had higher abundances of *Porphyromonas*, *Escherichia-Shigella*, *Enterococcus*, *Streptococcus*, and *Peptostreptococcus* genera (p <= 0.05, Mann-Whitney). The authors report that healthy controls were enriched *Bacteroides*, *Roseburia*, *Alistipes*, *Eubacterium*, and *Parasutterella* genera (p <= 0.05, Mann-Whitney). We found very similar results in our re-analysis of these data. We saw greater levels of *Enterococcus*, *Peptostreptococcus*, *Enterobacter*, *Klebsiella*, *Escherichia-Shigella*, and *Porphyromonas* genera in CRC patients (q <= 0.05, KW tests). And we observed significantly higher levels of *Bacteroides*, and several genera within *Lachnospiraceae* in healthy controls (q <= 0.05, KW tests). Furthermore, we also did not detect any significant differences in alpha diversity between CRC and healthy patients.

Chen et al. (2012) analyzed stool from 22 healthy patients and 21 CRC patients [10]. The authors found that *Paraprevotella*, *Eubacterium*, *Desulfovibrio*, *Mogibacterium*, *Collinsella*, *Anaerotruncus*, *Slackia*, *Anaerococcus*, *Porphyromonas*, *Fusobacterium*, and *Peptostreptococcus* genera were significantly enriched in CRC patients relative to controls, while *Bifidobacterium*, *Faecalibacterium*, and *Blautia* were reduced in CRC patients (p <= 0.05, Mann-Whitney). In our re-analysis of this data set, we found no significant differences between CRC and control patients. Again, this is likely due to the small number of replicates and our implementation of multiple-test corrections. However, non-significant trends were largely in agreement with the original results.

Across these four colorectal cancer studies, we find significant agreement. Dysbiosis associated with CRC is generally characterized by increased prevalence of *Fusobacterium*, *Porphyromonas*, *Peptostreptococcus*, *Parvimonas*, *Leptotrichia*, *Desulfovibrio*, and *Anaerococcus* genera (i.e. these genera were higher in CRC patients in 2 or more studies). In addition, there is a consistent decrease in the abundances of *Faecalibacterium*, *Blautia*, *Bacteroides* genera and organisms from the *Lachnospiraceae* family in CRC patients. CRC appears to have a smaller impact on overall community

structure than diahrrea. Indeed, we saw no significant differences in alpha diversity between healthy controls and CRC patients. In summary, CRC is characterized by a consistent enrichment of disease-associated bacteria.

Inflammatory Bowel Disease is characterized by a depletion of health-associated bacteria (IBD - Ulcerative Colitis and Crohn's Disease; 4 studies)

Gevers et al. (2014) looked for microbial signatures of Crohn's disease (CD) samples across 447 CD patients and 221 non-IBD controls [11]. Non-IBD controls were patients with non-inflammatory conditions such as abdominal pain and diarrhea. The authors report increased abundance of *Enterobacteriaceae*, *Pasteurellaceae*, *Veillonellaceae*, and *Fusobacteriaceae* in CD patients. CD patients also showed a drop in the abundances of *Erysipelotrichales*, *Bacteroidales*, and *Clostridiales* (*Ruminococcaceae* and *Lachnospiraceae*) taxa. These results were based on a mixture of 16S amplicon and shotgun metagenomic sequencing. In our re-analysis of the 16S stool data, we found significant enrichment in *Anaerosporobacter*, *Roseburia*, *Hespellia*, *Ruminococcus II*, *Eubacterium*, *Pseudoflavonifractor*, *Sporobacter*, *Ruminococcus*, *Subdoligranulum*, *Papillibacter*, *Collinsella*, and *Methanobrevibacter* in healthy patients (q <= 0.05, KW tests). The only genera that we saw significantly enriched in CD patients were *Lactobacillus* and *Acetanaerobacterium* (q <= 0.05, KW tests). We found a similar set of taxa enriched in the controls, but did not detect as many significant CD-enriched genera as the authors reported. This is likely due to the fact that we restricted our analysis to the 16S stool data. However, we saw non-significant trends in *Enterobacteriaceae* and *Veillonellaceae* consistent with the results reported in the original paper.

Morgan et al. (2012) studied a cohort of 119 CD patients, 74 UC patients, and 27 healthy controls [12]. The authors found that healthy patientsâĂŹ gut microbiomes were significantly enriched in *Roseburia*, *Phascolarctobacterium*, and an unclassified genus in the family *Veillonellaceae* (multivariate linear model, q <= 0.25). Patients with UC showed significantly higher levels of *Clostridiaceae* (multivariate

194

linear model, q <= 0.25). In our re-analysis, we did not find any genera that were significantly enriched in IBD patients. We found that healthy patients had significantly greater abundances of *Ruminococcus*, and *Gemmiger* relative to both UC and CD patients (q <= 0.05, KW tests). Additionally, CD patients were depleted in *Clostridium IV* relative to healthy controls (q <= 0.05, KW tests).

Papa et al. (2012) studied a cohort of 23 CD patients, 43 UC patients, and 24 non-IBD controls [13]. Non-IBD controls were patients with symptoms such as: constipation, abdominal pain, gastroesophageal reflux, poor weight gain, diarrhea, blood in stool and oropharyngeal dysphagia. At the genus level, they found that controls were enriched in *Alistipes, Subdoligranulum, Anaerovorax, Oscillibacter, Parabacteroides, Odoribacter, Ruminococcus, Butyricicoccus, Akkermansia, Anaerotruncus, Sporobacter, Phascolarctobacterium, Lawsonia, Ethanoligenens, Peptococcus* relative to IBD patients (KW, q < 0.01). The only genus that was found to be enriched in IBD patients was *Escherichia-Shigella*. In our re-analysis, we also found *Escherichia-Shigella* and *Cronobacter* to be enriched in patients with IBD (q <= 0.05, KW tests). When comparing healthy controls with UC patients, we also found an enrichment of *Haemophilus* in the UC patients. Control patients showed higher abundances of *Phascolarctobacterium, Butyricicoccus, Ruminococcus II, Oscillibacter, Ruminococcus, Gemmiger, Subdoligranulum, Clostridium IV, Odoribacter, Alistipes*, and *Parabacteroides* relative to all IBD patients (q <= 0.05, KW tests). Additionally, control patients were enriched in *Clostridium XIVa, Flavonifractor*, and *Akkermansia* relative to UC patients. Overall, our results match very closely what was found in the original paper.

Willing et al. (2010) compared 29 CD patients and 16 UC patients to 35 healthy controls [14]. The authors reported variable, and sometimes opposing shifts in the microbiomes of patients with UC, ileal CD and colonic CD at different taxonomic resolutions. We found no significant differences between IBD and healthy patients in our re-analysis. When comparing healthy controls with CD cases only, we found an enrichment of *Butyricicoccus* and *Oscillibacter* in the control patients (q <= 0.05, KW tests).

195

In summary, there are certain consistencies across IBD studies. IBD patients tend to be depleted in butyrate-producing clostridia: *Ruminococcus* and *Lachnospiraceae*. The organisms the are enriched in CD and UC patients tend to vary across studies. One consistency is organisms associated with the upper gut, like *Lactobacillus* and *Enterobacteriaceae* appear to be enriched in IBD patients [5]. This result fits with the reduced stool transit times associated with IBD (i.e. diarrhea).

Obesity shows a somewhat inconsistent microbial signature (OB; 5 studies)

Goodrich et al. (2014) studied a cohort of 416 twin pairs: 422 normal BMI, 322 overweight, and 185 obese [15]. The authors report higher levels of *Lactobacillaceae*, *Eggerthella*, and *Lachnospiraceae* (*Blautia* and *Dorea*) in obese individuals (q < 0.05, FDR-corrected T-test). They showed enrichment for *Christensenellaceae*, *Dehalobacterium*, *Lachnospira*, *Mogibacteriaceae*, *Rikenellaceae*, *Methanobre*, *Coriobacteriaceae*, *Peptococcaceae*, *Oscillospira*, *Ruminococcaceae*, and *Sarcina* in healthy BMI individuals (q < 0.05, FDR-corrected T-test). In our re-analysis, we found higher levels of *Streptococcus*, *Weissella*, *Roseburia*, *Blautia*, *Clostridium XlVb*, and *Mogibacterium* in obese individuals, while *Robinsoniella*, *Ruminococcaceae* (*Oscillibacter*, *Pseudoflavonifractor*, *Sporobacter*, and *Anaerofilum*), and *Anaerovorax* were more abundant in low-BMI individuals (q <= 0.05, KW tests). Our results only partially agree with the authors' original findings, which may be due to the fact that we used a different statistical test and OTU-calling method and that we binned the data at the genus level.

Zupancic et al. (2012) analyzed 310 individuals from an Amish population with varying BMIs [16]. They found a significant positive correlation between the abundance of *Collinsella* and BMI (i.e. enriched in obese individuals), while *Lachnobacterium*, *Anaerotruncus*, *Faecalibacterium*, and *Clostridium* were negatively correlated with BMI (i.e. enriched lean individuals) (p < 0.001, Spearman correlation). We found no significant differences in the proportion of genera between lean and obese individuals in our re-analysis.

Turnbaugh et al. (2008) looked differences in gut microbial community structure

between 31 monozygotic and 23 dizygotic twin pairs concordant for leanness or obesity [17]. The authors report a reduction in alpha diversity in obese individuals. They also report a significant decrease in *Bacteroidetes* and an increase in *Actinobacteria* in obese twins. In our re-analysis of these data, we did not see a significant reduction in alpha diversity (Supplementary Figure B-6). We found significant increases in *Catenibacterium, Acidaminococcus, Megasphaera, Lactobacillus, Roseburia*, and *Collinsella* in obese twins (q <= 0.05, KW tests). *Coprobacillus, Clostridium XVIII, Phascolarctobacterium, Clostridium XlVb, Oscillibacter, Flavonifractor, Pseudoflavonifractor, Ruminococcus, Clostridium IV, Gordonibacter, Alistipes*, and *Barnesiella* were significantly enriched in lean twins (q <= 0.05, KW tests).

Ross et al. (2015) looked at 63 Mexican American patients with varying BMIs [18]. They found no significant differences between patients with high and low BMIs within their 63 patient cohort, but identified several significant differences between their patient population and the HMP data set. However, it is unclear whether these differences were related to obesity, so we do not discuss them here. Our re-analysis of these results also found no significant differences in the relative abundances of bacterial genera between high- and low-BMI subjects.

Zhu et al. (2013) compared across a cohort of 16 healthy and 25 obese patients, in addition to 22 patients with Nonalcoholic steatohepatitis (see below) [19]. For obesity, the authors found that *Prevotella* was enriched in high-BMI patients, while healthy controls showed significantly greater relative abundances of *Bifidobacterium, Blautia*, and *Faecalibacterium* (p <= 0.05, ANOVA with post-hoc Tukey's tests). In our re-analysis of these data, we found a significant enrichment of *Peptoniphilus, Anaerococcus, Finegoldia, Leuconostoc, Mogibacterium, Varibaculum, Campylobacter, Prevotella*, and *Porphyromonas* in obese patients (q <= 0.05). Healthy patients were significantly enriched in *Akkermansia, Murdochiella, Blautia, Lachnospiracea incertae sedis*, and *Clostridium IV, Anaerovorax* (q <= 0.05, KW tests).

Overall, we found several differences between lean and obese patients that were consistent across at least two studies. *Roseburia* and *Mogibacterium* were enriched in obese individuals in more than one study. *Pseudoflavonifractor, Oscillobacter*,

Anaerovorax and *Clostridium IV* were enriched in the controls across more than one study. However, no genera showed consistent differences across three or more studies. Our results are largely consistent with a recent meta-analysis of obesity studies, which found no universal signature of human obesity [20].

Human Immunodeficiency Virus microbial signature is confounded with patient cohorts (HIV; 3 studies)

Dinh et al. (2015) compared the gut microbiome from 16 healthy patients to 22 patients with chronic HIV infections [21]. The authors report an general enrichment in *Proteobacteria* in HIV-infected patients. At the genus level, they found a significant enrichment in *Barnesiella* and a depletion in *Alistipes* in HIV-infected patients (LEfSe, $p < 0.05$). In our re-analysis of these data we found no significant differences in the relative abundances of genera between healthy and HIV-infected patients.

Lozupone et al. (2013) looked at 22 HIV-positive patients and 13 healthy controls [22]. The authors reported enrichment of *Prevotella, Catenibacterium, Dialister, Allisonella,* and *Megasphera* genera in HIV-positive patients, while *Bacteroides* and *Alistipes* were more abundant in controls ($p < 0.05$, ANOVA). We found all the associations reported above in our re-analysis. Additionally, we saw higher relative abundances of *Erysipelotrichaceae incertae sedis, Peptococcus, Mogibacterium, Peptostreptococcus, Desulfovibrio, Hallella,* and *Alloprevotella* in HIV-positive patients. And healthy patients were also enriched in *Oridibacter, Anaerostipes,* and *Parasutterella*. Many of the significant genera from the Lozupone study were shown to be strongly associated with sexual behavior in the Noguera-Julian study (i.e. these genera were significantly different in men who have sex with men versus other subjects; see below) and may not necessarily be related to HIV status.

Noguera-Julian et al. (2016) studied a cohort of 293 HIV-infected patients and 57 healthy controls. The authors found that many putative associations between HIV and the microbiome were driven by sexual preference (i.e. *Prevotella*, along with several other genera, were enriched in men who have sex with men (MSM)). After controlling for this demographic confounder, the authors reported that they were

not able to classify HIV positive and negative patients MSM patients. Due to the large size of their study, the authors were able to separate the influences of sexual behavior and HIV-status from one another and found that the majority of reported HIV-associations are likely confounded with sexual behavior.

Overall, there is not yet a strong consensus on the impacts of HIV on the human gut microbiome. Differences between patient cohorts may have obscured any putative HIV signal across studies. For example, all the patients in the Dinh et al. (2015) study were on antiretroviral therapy (ART), while only some of the patients in the other two studies were on ART. Noguera-Julian et al. (2016) found that patients who initiated ART within the first 6 months of HIV infection were able to maintain gut microbial community richness, unlike patients that were not on ART. In addition, the Noguera-Julian et al. (2016) paper was able to show that prior results showing enrichment of *Prevotella* in HIV-positive patients was an artifact due to this genera being enriched in men who have sex with men.

Autism Spectrum Disorder (ASD; 2 studies)

Kang et al. (2013) reported a reduced prevalence of *Prevotella* and other fermentative organisms in the guts of ASD children [23]. In particular, the authors showed significant (q <= 0.05, Mann-Whitney) depletion in unclassified *Prevotella* and *Veillonellaceae* genera in autistic children (n = 20 treatment and 20 controls). The authors also note a reduced alpha diversity in autistic children. After reprocessing these data, we found no significant differences in alpha diversity or genera abundances between autistic and control children (Figure 1; q > 0.05, Kruskal-Wallis). The original conclusion that *Prevotella* and *Veillonellaceae* were different was based on q-values of 0.04, which is only moderately convincing evidence against the null-hypothesis. Therefore, the loss of this marginal significance (for q <= 0.05) is unsurprising when using a different statistical test.

In a more recent study, Son et al. (2015) found no significant differences in microbial community diversity or composition between autistic and neurotypical children (n = 59 ASD and 44 neurotypical) [24]. One genus, representing chloroplast sequences,

was associated with ASD children with functional constipation, but this signal appeared to be due to dietary intake of chia seeds. Similar to the authorsâĂŹ findings, we did not detect any significant differences in genera abundances between ASD children and neurotypical children in the reprocessed data (q > 0.05, Kruskal-Wallis).

Taken together, we find no evidence for changes in the composition or diversity of the gut microbiome in response to ASD. However, we cannot discount subtle dysbiosis (i.e. small effect size) in response to ASD due to the small number of patients in each study.

Type 1 Diabetes (T1D; 2 studies)

Alkanani et al. (2015) compared 23 healthy patients to 35 early-onset T1D patients and 21 seropositive T1D patients [25]. The authors report higher relative abundances of *Lactobacillus*, *Prevotella* and *Staphylococcus* genera in healthy patients (p < 0.05, Wilcoxon). T1D patients showed higher levels of *Bacteroides* (p < 0.05, Wilcoxon). In our re-analysis, we found no significant differences in bacterial genera across healthy and diseased patients.

Mejia-Leon et al. (2014) compared 8 healthy patients to 8 early-onset T1D patients and 13 T1D patients who had received 2 years of treatment [26]. Similar to Alkanani et al. (2015), they found controls to be significantly enriched in *Prevotella* and T1D patients enriched in *Bacteroides* (p < 0.05, ANOVA, Tukey-Kramer test). They also found higher levels of *Acidaminococcus* and *Megamonas* genera (in the *Veillonellaceae* family) in the controls (p < 0.05, ANOVA, Tukey-Kramer test). We saw no significant differences in our re-analysis of these data.

Overall, the original authors report a consistent increase in *Bacteroides* and depletion in *Prevotella* genera associated with T1D. However, our re-analysis found that these differences did not pass our significance threshold. Thus, we cannot yet conclude that there is a consistent dysbiosis associated with T1D.

Nonalcoholic Steatohepatitis (NASH; 2 studies)

Zhu et al. (2013) compared the microbiomes from 16 healthy individuals to 22 patients with NASH [19]. They found significantly lower relative abundances of *Bifidobacterium*, *Blautia*, and *Faecalibacterium* genera in NASH patients (p <= 0.05, ANOVA with post-hoc Tukey's tests). NASH patients were enriched in *Escherichia*, compared to controls, and tended to show increased levels of *Proteobacteria* (p <= 0.05, ANOVA with post-hoc Tukey's tests). In our re-analysis, we found that NASH patients showed significantly higher levels of *Fusobacterium*, *Peptoniphilus*, *Anaerococcus*, *Finegoldia*, *Gallicola*, *Negativicoccus*, *Leuconostoc*, *Weissella*, *Lactobacillus*, *Peptococcus*, *Moryella*, *Syntrophococcus*, *Mogibacterium*, *Olsenella*, *Varibaculum*, *Mobiluncus*, *Pyramidobacter*, *Escherichia/Shigella*, *Campylobacter*, *Hallella*, *Prevotella*, and *Porphyromonas* genera (q < 0.05, KW test). Conversely, control patients were significantly enriched in *Akkermansia*, *Murdochiella*, *Coprococcus*, *Anaerostipes*, *Blautia*, *Lachnospiracea incertae sedis*, *Faecalibacterium*, *Ruminococcus*, *Gemmiger*, *Clostridium IV*, *Anaerovorax*, *Clostridium XI*, *Corynebacterium*, *Bifidobacterium*, *Alistipes*, and *Barnesiella* genera (q < 0.05, KW test).

Wong et al. (2013) investigated a cohort of 16 healthy and 22 NASH patients [27]. They found that control patients were enriched in *Faecalibacterium* and *Anaerosporobacter* genera, while NASH patients showed significantly higher levels of *Parabacteroides* and *Alisonella* genera (p < 0.05, t-test). In our re-analysis of these data, we saw no significant differences.

In summary, there were not many consistencies between the two NASH studies analyzed here. The original studies consistently report a depletion in *Faecalibacterium* in NASH patients. Thus, the overall influence of NASH on the microbiome is difficult to assess without further study.

Minimal Hepatic Encephalopathy and Liver Cirrhosis (LIV; 1 study)

Zhang et al. (2013) looked at the microbiomes of 26 healthy patients, 26 patients with MHE, and 25 patients with CIRR [28]. The original paper reported several gen-

era that differed between diseased and control patients. *Odoribacter*, *Flavonifractor*, and *Coprobacillus* were all enriched in MHE patients relative to controls, while *Eubacterium*, *Lachnospira*, *Parasutteralla*, and an unclassified *Erysipelotrichaceae* genus were enriched in healthy patients ($p < 0.01$, Mann-Whitney). The authors also reported depletion in *Prevotella* in non-MHE patients with cirrhosis (CIRR), relative to controls. When we re-processed and re-analyzed these data, the only difference we found was an enrichment in *Veillonella* in case (MHE and CIRR) patients ($q < 0.05$, KW test). When comparing controls with MHE patients alone, we also saw an enrichment of *Faecalibacterium* in healthy controls relative to MHE cases.

Rheumatoid and Psoriatic Arthritis (ART; 1 study)

Scher et al. (2013) investigated the impacts of arthritis on a cohort of 86 arthritic and 28 healthy patients [29]. The authors report that greater abundances of *Prevotella copri* can predict susceptibility to arthritis. There were three types of arthritic conditions studied, but only new-onset untreated rheumatoid arthritis (NORA) showed a strong association with multiple *Prevotella* OTUs among others ($q < 0.01$, LEfSe). The other RA groups were not easily distinguishable from controls. Indeed, when grouping all arthritis patients together for our re-analysis as well as comparing RA and psoriatic arthritis patients separately, we did not find any genera that were significantly different between arthritic patients and controls.

Parkinson's Disease (PAR; 1 study)

Scheperjans et al. (2014) looked for differences in the gut microbiome between 72 neurotypical patients and 72 Parkinson's (PAR) patients [30]. They found a small handful of significant differences at the family level. Control patients showed higher relative abundances of *Prevotellaceae*, while PAR patients were enriched in *Lactobacillaceae*, *Verrucomicrobiaceae*, *Bradyrhizobiaceae*, and *Clostridiales Incertae Sedis* ($q < 0.05$, Mann-Whitney). In our re-analysis, we found significantly higher relative abundances of *Lactobacillus* (within *Lactobacillaceae*) and *Alistipes* (within *Rikenellaceae*) in PAR patients ($q < 0.05$, KW tests).

B.1.2 Stratifying heterogenous case groups shows consistent disease-specific signals

In our main analyses, we combined Crohn's disease (CD) and ulcerative colitis (UC) patients together as IBD cases. We also performed separate analyses on these individual patient groups. All four IBD studies included CD cases and three included UC cases (all except Gevers et al. (2014) [11]). We performed the same analysis as in Figure 1 for these stratified groups, and found that both CD and UC patients are characterized by depletion of similar health-associated microbes (Supplementary Figures B-4 and B-5). Interestingly, neither UC nor CD seemed to have a larger microbiome shift: only one dataset for each type of comparison had more than 10 significant genera (Gevers et al. (2014), 14 CD-associated genera; Papa et al. (2012), 17 UC-associated genera). Additional studies comparing UC- and CD-specific microbiome alterations will be needed to tease out whether and how these IBD subtypes differentially impact the gut microbiome.

We also performed stratified analyses on the arthritis (ART) and liver (LIV) patients in the Scher et al. (2013) and Zhang et al. (2013) datasets, respectively [28, 29] (Supplementary Figure B-4). The random forest classifiers performed similarly well on the stratified patient groups than on the combined cases. As in the combined analyses, neither type of arthritis (rheumatoid arthritis (RA) or psoriatic arthritis (PSA)) had any significant genus-level associations. In the Zhang et al. (2013) dataset, 1 genus was significantly associated with the liver cirrhosis (CIRR) patients and 2 with the minimal hepatic encelopathy (MHE) patients. As in the original combined analysis, *Veillonella* was associated with both groups of patients. In our stratified analysis, *Faecalibacterium* was additionally significantly associated with non-MHE healthy controls. However, the lack of other arthritis or liver datasets in this analysis prevents us from drawing more generalized conclusions from these stratified analyses.

B.1.3 Healthy vs. disease classifier identifies general microbiome shifts

To further address the question of whether we could find a robust, generalized signal for diseased microbiomes regardless of the disease type, we built two classifiers to distinguish healthy patients from any type of case patients. In these classifiers, we excluded the two datasets which did not have healthy controls (Gevers et al. (2014) [11] and Papa et al. (2012) [13], which used non-IBD patients as controls) and CDI Youngster (2014), [3] which had only 4 distinct controls. First, we performed leave-one-dataset-out cross-validation to determine whether a general healthy vs. disease classifier trained on the other datasets could still classify cases from controls in a test dataset. These AUCs correlated well with the single-dataset classifiers, though usually performed slightly less well than the single-dataset classifiers (Pearson $\rho = 0.56$, p = 0.003; Supplementary Figure B-9). We also built a more stringent leave-one-disease-out classifier to ensure that the diarrhea datasets and others with strong microbiome signals were not driving the classification ability of all other diseases. Surprisingly, this classifier performed similarly to the leave-one-dataset-out classifier (Supplementary Figure B-9). The positive correlation with the original single-dataset classification results (Pearson $\rho = 0.47$, p = 0.02) indicates that there is a generalizable healthy vs. disease microbiome signal that is being identified even across different diseases. These results also indicate that models for each disease group are predictive of cases and controls for other datasets within that group, since the leave-one-dataset-out classifier, which included datasets of the test disease group in the training set, performed better than the leave-one-disease-out classifier, which did not.

B.1.4 Shared microbial response is robust to different definitions

Our simple heuristic defined non-specific microbes as those which were significantly enriched or depleted in two diseases. To ensure that this definition was not being dominated by the diarrhea datasets and that we were indeed identifying microbes

which respond non-specifically to multiple diseases, we re-defined the non-specific genera as those which were significantly enriched or depleted in two diseases, excluding datasets with diarrhea cases (Schubert et al. (2014) [1], Singh et al. (2015) [4], Vincent et al. (2013) [2], and Youngster et al. (2014) [3]). We found that 27 out of the 51 original non-specific genera were recovered, with all health- and disease-associated effects in matching directions (Supplementary Figure B-10). Thus, the majority of the shared microbial response is robust to the exclusion of diarrhea datasets.

We also re-defined non-specific microbes using Stouffer's method to combine p-values across all datasets (except Papa et al. (2012) [13], Gevers et al. (2014) [11], and Lozupone et. al (2013) [22]) [31]. We combined each dataset's FDR-corrected q-values with `scipy.stats.combine_pvalues(method='stouffer')`, using the square root of each study's sample size as the weights. Genera with a combined q-value less than 0.05 were considered non-specific responders. Overall, these results did not conflict with the heuristic definition (i.e. only two genera, *Porphyromonas* and *Gemmiger*, were "health-associated" with one method and "disease-associated" with the other; Supplementary Figure B-10). Stouffer's method is less conservative than the heuristic definition, identifying 111 genera in the non-specific response (60 health-associated and 51 disease-associated). In addition, using Stouffer's method does not allow for the identification of mixed genera (i.e. those which respond in both health- and disease-associated directions across multiple diseases). Finally, combining q-values with Stouffer's method does not ensure that identified microbes are responding non-specifically to multiple diseases: one highly significant genus in a large study can dominate other q-values and be flagged as a non-specific responder, despite only being associated with one disease. Thus, the heuristic definition is more conservative and more directly related to the biological question of identifying shared microbial responses to disease.

We tested whether the overall number of non-specific responders we observed was greater than we would expect to see due to chance. We built an empirical null distribution of the number of each type of non-specific responder. We shuffled q-values within each dataset, re-defined non-specific responders, and counted how

many health-associated, disease-associated, and mixed genera were found, repeating this process 1000 times. When we considered significance in two diseases as the threshold for our heuristic (as presented in the main text), we did not find a significantly larger number of non-specific responses than would be expected by chance (Supplementary Figure B-11). When we raised the heuristic threshold to three diseases our results became more significant, but there was a large reduction in the number of identified non-specific genera. Thus, there is currently not enough information to fully distinguish between microbes that are sporadically detected across multiple diseases from those that may be consistently associated with general health or disease. Future meta-analyses that include many more datasets for each of many conditions might be able to distinguish microbes that are consistently associated with health or disease from those that are sporadically associated with different conditions.

Despite the fact that the number of non-specific microbes did not reach statistical significance, we identified multiple lines of evidence for a coherent microbial response to health and disease. First, the healthy vs. disease classifiers successfully classified case patients across a variety of diseases even when the disease being tested was not in the training set, indicating that some aspects of disease-associated microbiome shifts can generalize across diseases (Supplementary Figure B-9). Second, the statistical significance of the number of non-specific responders increased as we increased the number of diseases threshold (Supplementary Figure B-11). Thus, future meta-analyses which include many more studies and disease states may be able to more robustly identify bacteria which respond across a broader variety of disease states. Third, we saw a coherent phylogenetic signal in the non-specific response (e.g. Proteobacteria and *Lactobacillaceae* associated with disease and *Rumminococcaceae* and *Lachnospiraceae* associated with health), which points to potential mechanisms (e.g. shorter stool transit time or inflammation) for a shared response to health or disease (Figure 3A). Thus, we expect that future meta-analyses that include more studies and diseases will identify a consistent set of bacteria that form a general microbial response to health and disease in the gut.

B.2   Supplementary Tables

Dataset ID	Year	Controls	N (controls)	Cases	N (cases)	Median reads per sample	Sequencer	16S Region	Ref.
Scher 2013, ART	2013	H	28	PSA, RA	86	2194.0	454	V1-V2	[29]
Kang 2013, ASD	2013	H	20	ASD	19	1345.0	454	V2-V3	[23]
Son 2015, ASD	2015	H	44	ASD	59	4777.0	Miseq	V1-V2	[24]
Schubert 2014, CDI	2014	H	154	CDI	93	4897.0	454	V3-V5	[1]
Schubert 2014, nonCDI	2014	H	154	nonCDI	89	4903.0	454	V3-V5	[1]
Singh 2015, EDD	2015	H	82	EDD	201	2585.0	454	V3-V5	[4]
Vincent 2013, CDI	2013	H	25	CDI	25	2526.5	454	V3-V5	[2]
Youngster 2014, CDI	2014	H	4	CDI	19	15081.0	Miseq	V4	[3]
Baxter 2016, CRC	2016	H	172	CRC	120	9913.5	Miseq	V4	[6]
Chen 2012, CRC	2012	H	22	CRC	21	1152.0	454	V1-V3	[10]
Wang 2012, CRC	2012	H	54	CRC	44	161.0	454	V3	[9]
Zeller 2014, CRC	2014	H	75	CRC	41	120989.0	MiSeq	V4	[8]
Dinh 2015, HIV	2015	H	15	HIV	21	3248.5	454	V3-V5	[21]
Lozupone 2013, HIV	2013	H	13	HIV	23	3262.0	MiSeq	V4	[22]
Noguera-Julian 2016, HIV	2016	H	34	HIV	205	16506.0	MiSeq	V3-V4	[32]
Gevers 2014, IBD	2014	nonIBD	16	CD	146	9773.5	Miseq	V4	[11]
Morgan 2012, IBD	2012	H	18	UC, CD	108	1022.5	454	V3-V5	[12]
Papa 2012, IBD	2012	nonIBD	24	UC, CD	66	1323.5	454	V3-V5	[13]
Willing 2010, IBD	2009	H	35	UC, CD	45	1118.5	454	V5-V6	[14]
Zhang 2013, LIV	2013	H	25	CIRR, MHE	46	487.0	454	V1-V2	[28]
Wong 2013, NASH	2013	H	22	NASH	16	1980.0	454	V1-V2	[27]
Zhu 2013, NASH	2013	H	16	NASH	22	10863.0	454	V4	[19]
Goodrich 2014, OB	2014	H	428	OB	185	27077.0	Miseq	V4	[15]
Ross 2015, OB	2015	H	26	OB	37	4562.0	454	V1-V3	[18]
Turnbaugh 2009, OB	2009	H	61	OB	195	1556.5	454	V2	[17]
Zhu 2013, OB	2013	H	16	OB	25	9778.0	454	V4	[19]
Zupancic 2012, OB	2012	H	96	OB	101	1645.0	454	V1-V3	[16]
Scheperjans 2015, PAR	2015	H	74	PAR	74	2351.5	454	V1-V3	[30]
Alkanani 2015, T1D	2015	H	55	T1D	57	9117.0	MiSeq	V4	[25]
Mejia-Leon 2014, T1D	2014	H	8	T1D	21	4702.0	454	V4	[26]

Table B.1: Datasets collected and processed through standardized pipeline. Disease labels: ART = arthritis, ASD = austism spectrum disorder, CD = Crohn's disease, CDI = *Clostridium difficile* infection, CIRR = liver cirrhosis, CRC = colorectal cancer, EDD = enteric diarrheal disease, H = healthy, HIV = human immunodeficiency virus, LIV = liver diseases, MHE = minimal hepatic encephalopathy, NASH = non-alcoholic steatohepatitis, OB = obesity, PAR = Parkinson's disease, PSA = psoriatic arthritis, RA = rheumatoid arthritis, T1D = type I diabetes, UC = ulcerative colitis. nonCDI controls are patients with diarrhea who tested negative for *C. difficile* infection. nonIBD controls are patients with gastrointestinal symptoms but no intestinal inflammation. Datasets are ordered alphabetically by disease and within disease by first author.

Dataset ID	Data type	Barcodes	Primers	Quality filtering	Quality cutoff	Length trim
Scher 2013, ART	fastq	No	Yes	-fastq_truncqual	25	200
Kang 2013, ASD	fastq	No	Yes	-fastq_truncqual	25	200
Son 2015, ASD	fastq	No	Yes	-fastq_truncqual	25	200
Schubert 2014, CDI	fastq	No	Yes	-fastq_truncqual	25	150
Vincent 2013, CDI	fastq	No	Yes	-fastq_truncqual	20	101
Youngster 2014, CDI	fastq	No	No	-fastq_truncqual	25	200
Baxter 2016, CRC	fastq	No	No	-fastq_truncqual	25	250
Chen 2012, CRC	fastq	Yes	Yes	-fastq_truncqual	25	200
Wang 2012, CRC	fastq	Yes	Yes	-fastq_truncqual	25	150
Zeller 2014, CRC	fastq	No	No	-fastq_truncqual	25	200
Singh 2015, EDD	fasta	n/a	n/a	n/a	n/a	200
Dinh 2015, HIV	fastq	No	No	-fastq_truncqual	25	200
Lozupone 2013, HIV	fastq	No	No	-fastq_truncqual	25	150
Noguera-Julian 2016, HIV	fastq	No	Yes	-fastq_truncqual	25	200
Gevers 2014, IBD	fastq	No	No	-fastq_truncqual	25	200
Morgan 2012, IBD	fastq	No	Yes	-fastq_truncqual	25	200
Papa 2012, IBD	fasta	n/a	n/a	n/a	n/a	200
Willing 2010, IBD	fastq	No	Yes	-fastq_maxee	2	200
Zhang 2013, LIV	fastq	No	Yes	-fastq_truncqual	25	200
Wong 2013, NASH	fastq	No	No	-fastq_truncqual	25	200
Zhu 2013, NASH	fasta	n/a	n/a	n/a	n/a	200
Schubert 2014, nonCDI	fastq	No	Yes	-fastq_truncqual	25	150
Goodrich 2014, OB	fastq	No	No	-fastq_truncqual	25	200
Ross 2015, OB	fastq	No	No	-fastq_truncqual	25	150
Turnbaugh 2009, OB	fasta	n/a	n/a	n/a	n/a	200
Zhu 2013, OB	fasta	n/a	n/a	n/a	n/a	200
Zupancic 2012, OB	fastq	No	No	-fastq_truncqual	25	200
Scheperjans 2015, PAR	fastq	No	Yes	-fastq_truncqual	25	200
Alkanani 2015, T1D	fastq	No	No	-fastq_maxee	2	200
Mejia-Leon 2014, T1D	fastq	Yes	Yes	-fastq_truncqual	25	150

Table B.2: Processing parameters for all datasets. Barcodes column indicates whether we assigned reads to samples by their barcodes (Yes) or if the files were already de-multiplexed (No). Primers column indicates whether we removed the primers from sequences. Quality filtering and Quality cutoff columns indicate the type of quality filtering we performed on the data. Length trim is the length to which all sequences were truncated before clustering into OTUs. In the case of -fastq_truncqual quality filtering, reads were length trimmed after quality truncation. In the case of -fastq_maxee quality filtering, reads were length trimmed before quality filtering. Datasets are ordered alphabetically by disease and within disease by first author. ART = arthritis, ASD = autism spectrum disorder, CDI = *Clostridium difficile* infection, CRC = colorectal cancer, EDD = enteric diarrheal disease, HIV = human immunodeficient virus, IBD = inflammatory bowel disease, LIV = liver disease, NASH = non-alcoholic steatohepatitis, nonCDI = non-*Clostridium difficile* infection, OB = obesity, PAR = Parkinson's disease, T1D = type I diabetes.

Dataset ID	Raw data	Metadata
Scher 2013, ART	SRA study SRP023463	SRA
Kang 2013, ASD	SRA study SRP017161	SRA
Son 2015, ASD	SRA study SRP057700	SRA
Schubert 2014, CDI	mothur.org/CDI_MicrobiomeModeling	mothur.org/CDI_MicrobiomeModeling
Vincent 2013, CDI	email authors	email authors
Youngster 2014, CDI	SRA study SRP040146	email authors
Baxter 2016, CRC	SRA study SRP062005	SRA
Chen 2012, CRC	SRA study SRP009633	SRA sample description
Wang 2012, CRC	SRA study SRP005150	SRA study description
Zeller 2014, CRC	ENA study PRJEB6070	Table S1 and S2
Singh 2015, EDD	http://dx.doi.org/10.6084/m9.figshare.1447256	Additional File 4
Dinh 2015, HIV	SRA study SRP039076	SRA
Lozupone 2013, HIV	ENA study PRJEB4335	Qiita study 1700
Noguera-Julian 2016, HIV	SRA study SRP068240	SRA
Gevers 2014, IBD	SRA study SRP040765	Table S2
Morgan 2012, IBD	SRA study SRP015953	http://huttenhower.sph.harvard.edu/ibd2012
Papa 2012, IBD	email authors	email authors
Willing 2010, IBD	email authors	email authors
Zhang 2013, LIV	SRA study SRP015698	SRA
Wong 2013, NASH	SRA study SRP011160	SRA
Zhu 2013, NASH	MG-RAST, study mgp1195	MG-RAST
Schubert 2014, nonCDI	mothur.org/CDI_MicrobiomeModeling	mothur.org/CDI_MicrobiomeModeling
Goodrich 2014, OB	ENA studies PRJEB6702 and PRJEB6705	ENA
Ross 2015, OB	SRA study SRP053023	SRA
Turnbaugh 2009, OB	https://gordonlab.wustl.edu/NatureTwins_2008/TurnbaughNature_11_30_08.html	Table S1
Zhu 2013, OB	MG-RAST, study mgp1195 (same data as nash_zhu)	MG-RAST
Zupancic 2012, OB	SRA study SRP002465	SRA
Scheperjans 2015, PAR	ENA study PRJEB4927	sample names
Alkanani 2015, T1D	email authors	email authors
Mejia-Leon 2014, T1D	email authors	email authors

Table B.3: Locations of raw data and associated metadata for each dataset used in these analyses. Datasets are ordered alphabetically by disease and within disease by first author. ART = arthritis, ASD = autism spectrum disorder, CDI = *Clostridium difficile* infection, CRC = colorectal cancer, EDD = enteric diarrheal disease, HIV = human immunodeficient virus, IBD = inflammatory bowel disease, LIV = liver disease, NASH = non-alcoholic steatohepatitis, nonCDI = non-*Clostridium difficile* infection, OB = obesity, PAR = Parkinson's disease, T1D = type I diabetes.

Dataset ID	AUC	Fisher's p	Kappa score
Singh 2015, EDD	0.96	7.9e-31	0.7
Schubert 2014, CDI	0.99	8.7e-49	0.88
Schubert 2014, nonCDI	0.98	6.3e-38	0.79
Vincent 2013, CDI	0.91	1.6e-06	0.68
Goodrich 2014, OB	0.67	0.00014	0.11
Turnbaugh 2009, OB	0.84	1.7e-06	0.28
Zupancic 2012, OB	0.44	0.16	-0.11
Ross 2015, OB	0.49	0.75	-0.068
Zhu 2013, OB	0.86	1.3e-05	0.69
Baxter 2016, CRC	0.77	5.4e-16	0.43
Zeller 2014, CRC	0.82	3.4e-06	0.41
Wang 2012, CRC	0.9	2.6e-11	0.67
Chen 2012, CRC	0.78	0.034	0.35
Gevers 2014, IBD	0.71	1	0
Morgan 2012, IBD	0.81	0.0025	0.26
Papa 2012, IBD	0.84	0.0019	0.34
Willing 2010, IBD	0.66	0.81	0.026
Noguera-Julian 2016, HIV	0.67	1	0
Lozupone 2013, HIV	0.92	8.7e-06	0.76
Dinh 2015, HIV	0.22	0.062	-0.26
Son 2015, ASD	0.39	0.12	-0.16
Kang 2013, ASD	0.76	0.056	0.33
Alkanani 2015, T1D	0.71	0.0078	0.27
Mejia-Leon 2014, T1D	0.77	0.18	0.25
Wong 2013, NASH	0.68	0.098	0.28
Zhu 2013, NASH	0.93	1.3e-07	0.84
Scher 2013, ART	0.62	1	-0.034
Zhang 2013, LIV	0.8	0.016	0.29
Scheperjans 2015, PAR	0.67	0.0083	0.23

Table B.4: Area under the ROC curve (AUC), Fisher's p-values, and Kappa score for each case vs. control classifier. Metrics were calculated from the predictions on each test set in five-fold cross-validation. Datasets are ordered as in Figure 1. ART = arthritis, ASD = autism spectrum disorder, CDI = *Clostridium difficile* infection, CRC = colorectal cancer, EDD = enteric diarrheal disease, HIV = human immunodeficient virus, IBD = inflammatory bowel disease, LIV = liver disease, NASH = non-alcoholic steatohepatitis, nonCDI = non-*Clostridium difficile* infection, OB = obesity, PAR = Parkinson's disease, T1D = type I diabetes.
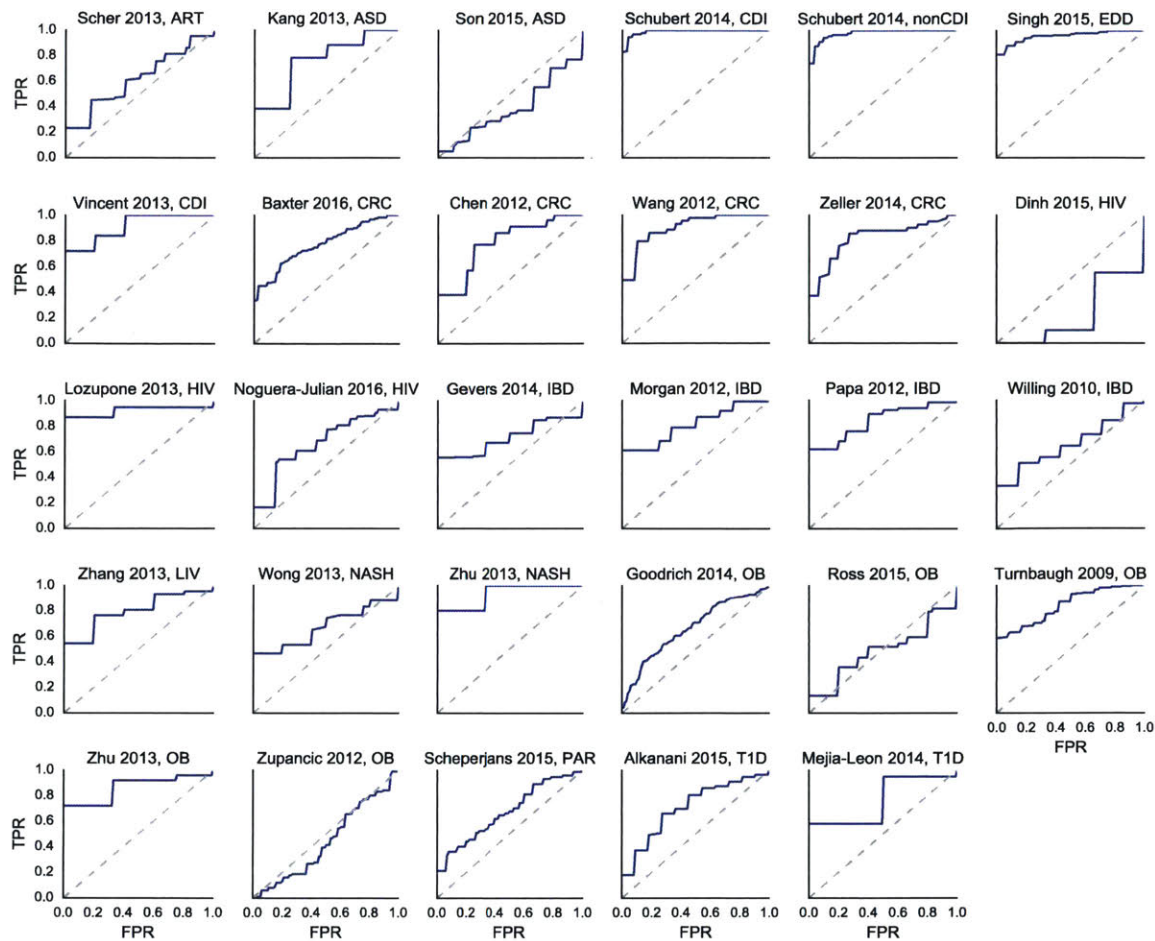
B.3 Supplementary Figures

Figure B-1: ROC curves for each of the classifiers in Figure 1. Datasets are ordered alphabetically by disease and within disease by first author. FPR = false positive rate, TPR = true positive rate. ART = arthritis, ASD = autism spectrum disorder, CDI = *Clostridium difficile* infection, CRC = colorectal cancer, EDD = enteric diarrheal disease, HIV = human immunodeficient virus, IBD = inflammatory bowel disease, LIV = liver disease, NASH = non-alcoholic steatohepatitis, nonCDI = non-*Clostridium difficile* infection, OB = obesity, PAR = Parkinson's disease, T1D = type I diabetes.

212
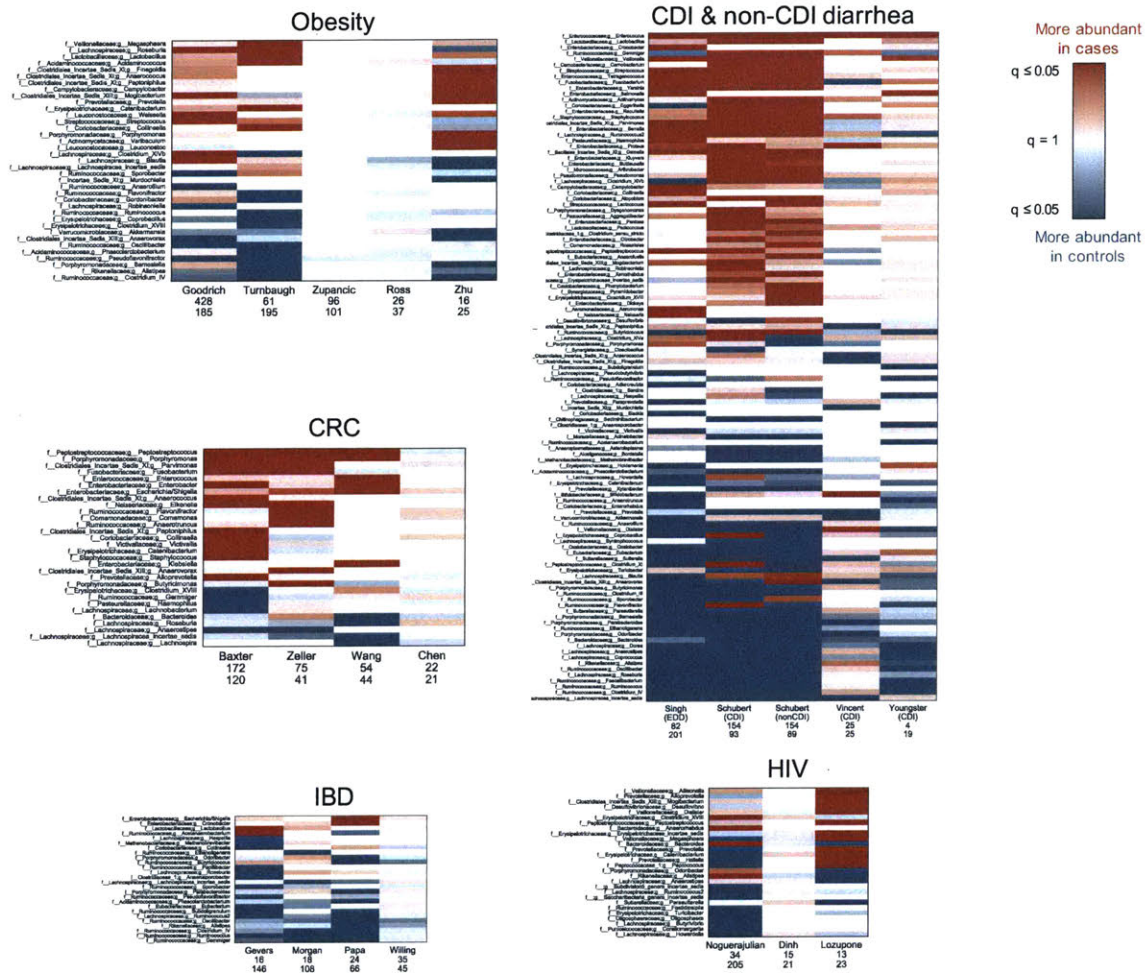
Figure B-2: Same heatmaps as in Figure 2, with rows labeled by family and genus taxonomy. Heatmaps show log10(q-values) for each disease (Kruskal-Wallis (KW) test, Benjamini-Hochberg FDR correction). Rows include all genera which were significant in at least one dataset within each disease, columns are datasets. Q-values are colored by direction of the effect, where red indicates higher mean abundance in disease patients and blue indicates higher mean abundance in controls. Opacity ranges from q = 0.05 to 1, where q values less than 0.05 are the most opaque and q values close to 1 are gray. White indicates that the genus was not present in that dataset. Within each heatmap, rows are ordered from most disease-associated (top) to most health-associated (bottom) (i.e. by the sum across rows of the log10(q-values), signed according to directionality of the effect).

213

Figure B-3: Panel A from Figure 3, with genus labels. Non-specific and disease-associated genera. Genera are in rows, arranged phylogenetically according to a PhyloT tree built from genus-level NCBI IDs (http://phylot.biobyte.de). Non-specific genera are associated with health (or disease) in at least two different *diseases* (q < 0.05, Kruskal-Wallis (KW) test, Benjamini-Hochberg FDR correction). Disease-specific genera are significant in the same direction in at least two *studies* of the same disease (q < 0.05, FDR KW test). As in Figure 2, blue indicates higher mean abundance in controls and red indicates higher mean abundance in patients. Black bars indicate mixed genera which were associated with health in two diseases and also associated with disease in two diseases. Disease-specific genera are shown for diseases with at least 3 studies.
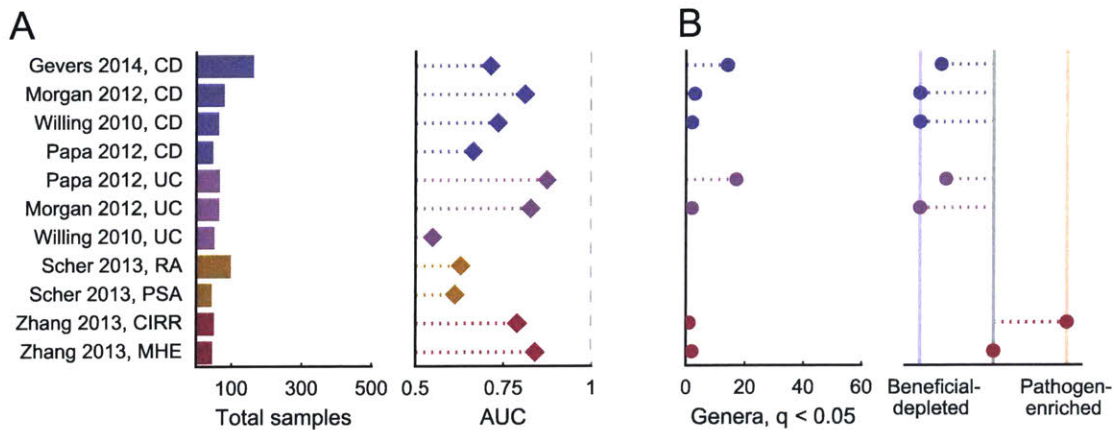
Figure B-4: Same analysis as in Figure 1 for stratified patient groups. (A) Left: Total sample size for each comparison. Right: Area under the ROC curve (AUC) for genus-level random forest classifiers. (B) Left: Number of genera with q < 0.05 (Kruskal-Wallis (KW) test, Benjamini-Hochberg FDR correction) for each type of patient group comparison. Right: Direction of microbiome shift,i.e. the percent of total associated genera which were enriched in diseased patients. In comparisons on the leftmost blue line, 100% of associated (q < 0.05, FDR KW test) genera are health-associated (i.e. depleted in patients relative to controls). In comparisons on the rightmost red line, 100% of associated (q < 0.05, FDR KW test) genera are disease-associated (i.e. enriched in patients relative to controls).

UC

	Papa	Morgan	Willing
	24	18	35
	43	47	16

CD

	Gevers	Morgan	Willing	Papa
	16	18	35	24
	146	61	29	23
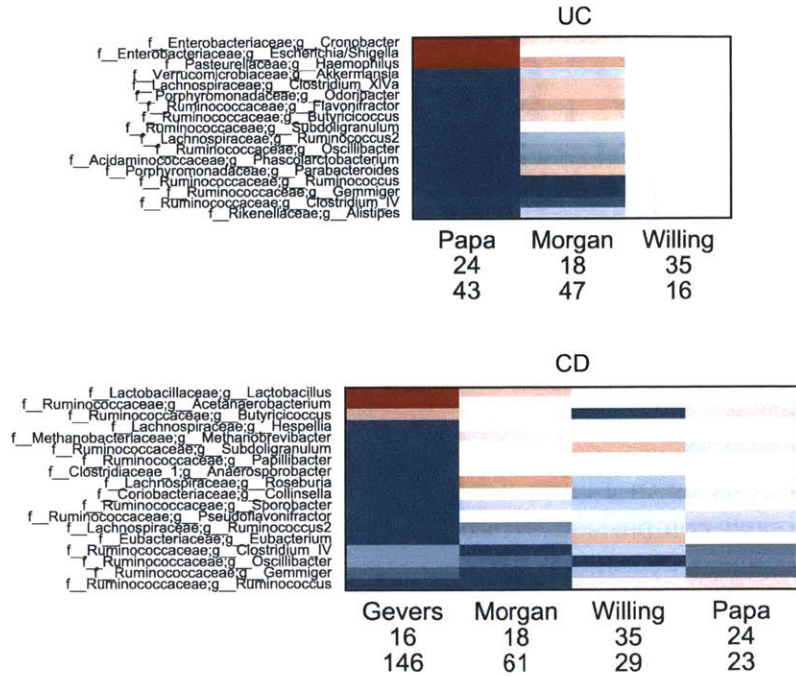
Figure B-5: Same results as presented in Figure 2 for ulcerative colitis (UC) and Crohn's disease (CD) IBD patients separately. Heatmaps show log10(q-values) for each comparison, with studies in columns and genera in rows (Kruskal-Wallis (KW) test, Benjamini-Hochberg FDR correction). Q-values are colored by direction of the effect, where red indicates higher mean abundance in disease patients and blue indicates higher mean abundance in controls. Opacity ranges from q = 0.05 to 1, where q values less than 0.05 are the most opaque and q values close to 1 are gray. White indicates that the genus was not present in that dataset. Within each heatmap, rows are ordered from most disease-associated (top) to most health-associated (bottom) (i.e. by the sum across rows of the log10(q-values), signed according to directionality of the effect).
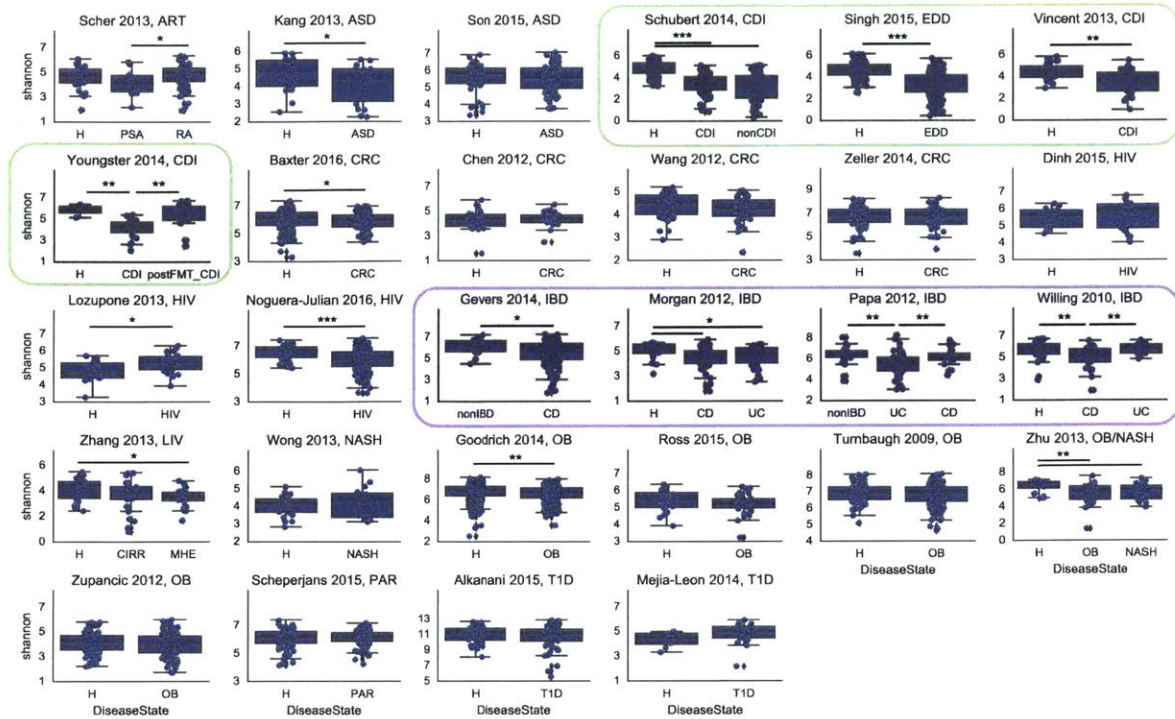
Figure B-6: Reduction in alpha diversity is not a reliable indicator of "dysbiosis." Shannon alpha diversity index across all patient groups in all studies, calculated on OTUs (i.e. not collapsed to genus level, and including unannotated OTUs). Diarrheal patients consistently have lower alpha diversity than non-diarrheal controls (green box). Crohn's disease (CD) patients also show a slight reduction of alpha diversity relative to controls in three out of four IBD studies and ulcerative colitis (UC) patients in two studies (purple box). Obese patients have inconsistent and small reductions in alpha diversity, consistent with a previous meta-analysis [20]. $* : 0.01 < p < 0.05, ** : 10^{-4} < p < 0.01, *** : p < 10^{-4}$. P values are calculated from a two-sided T-test (using `scipy.stats.ttest_ind`) and are not corrected for multiple tests. Note that the datasets with multiple case groups (Zhu et al. (OB/NASH, 2013) and Schubert et al. (CDI/non-CDI, 2014)) are presented only once in this plot. ART = arthritis, ASD = austism spectrum disorder, CD = Crohn's disease, CDI = *Clostridium difficile* infection, CIRR = liver cirrhosis, CRC = colorectal cancer, EDD = enteric diarrheal disease, H = healthy, HIV = human immunodeficiency virus, LIV = liver diseases, MHE = minimal hepatic encephalopathy, NASH = non-alcoholic steatohepatitis, OB = obesity, PAR = Parkinson's disease, PSA = psoriatic arthritis, RA = rheumatoid arthritis, T1D = type I diabetes, UC = ulcerative colitis. nonCDI controls are patients with diarrhea who tested negative for *C. difficile* infection. nonIBD controls are patients with gastrointestinal symptoms but no intestinal inflammation.
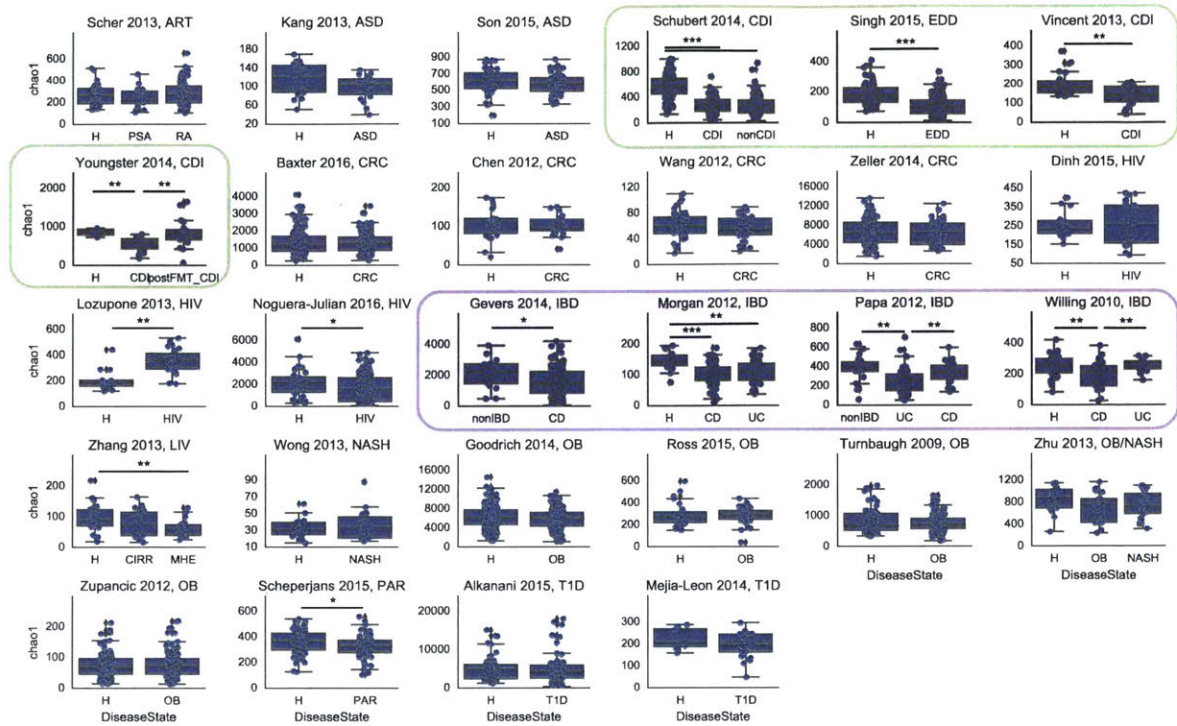
217

Figure B-7: Chao1 alpha diversity across all patient groups in all studies, calculated on OTUs (i.e. not collapsed to genus level, and including unannotated OTUs). $*: 0.01 < p < 0.05, **: 10^{-4} < p < 0.01, ***: p < 10^{-4}$. P values are calculated from a two-sided T-test (using `scipy.stats.ttest_ind`) and are not corrected for multiple tests. Note that the datasets with multiple case groups (Zhu et al. (OB/NASH, 2013) and Schubert et al. (CDI/non-CDI, 2014)) are presented only once in this plot. ART = arthritis, ASD = austism spectrum disorder, CD = Crohn's disease, CDI = *Clostridium difficile* infection, CIRR = liver cirrhosis, CRC = colorectal cancer, EDD = enteric diarrheal disease, H = healthy, HIV = human immunodeficiency virus, LIV = liver diseases, MHE = minimal hepatic encephalopathy, NASH = non-alcoholic steatohepatitis, OB = obesity, PAR = Parkinson's disease, PSA = psoriatic arthritis, RA = rheumatoid arthritis, T1D = type I diabetes, UC = ulcerative colitis. nonCDI controls are patients with diarrhea who tested negative for *C. difficile* infection. nonIBD controls are patients with gastrointestinal symptoms but no intestinal inflammation.
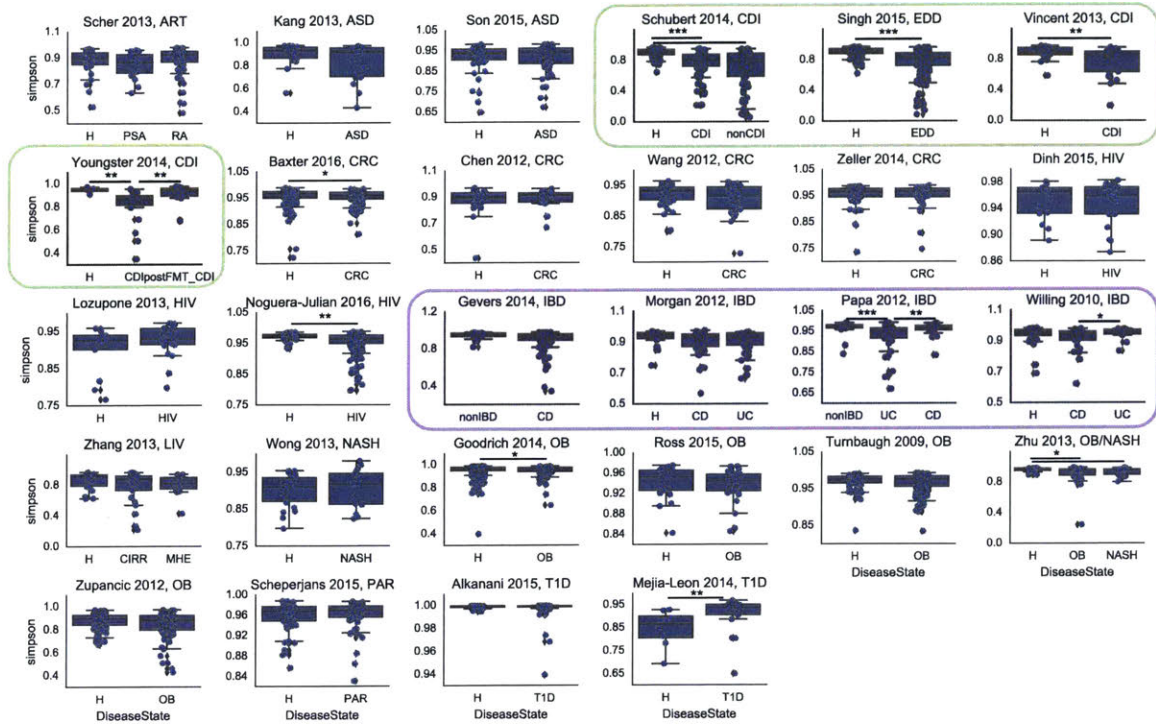
Figure B-8: Simpson alpha diversity across all patient groups in all studies, calculated on OTUs (i.e. not collapsed to genus level, and including unannotated OTUs). $*: 0.01 < p < 0.05, ** : 10^{-4} < p < 0.01, *** : p < 10^{-4}$. P values are calculated from a two-sided T-test (using `scipy.stats.ttest_ind`) and are not corrected for multiple tests. Note that the datasets with multiple case groups (Zhu et al. (OB/NASH, 2013) and Schubert et al. (CDI/non-CDI, 2014)) are presented only once in this plot. ART = arthritis, ASD = austism spectrum disorder, CD = Crohn's disease, CDI = *Clostridium difficile* infection, CIRR = liver cirrhosis, CRC = colorectal cancer, EDD = enteric diarrheal disease, H = healthy, HIV = human immunodeficiency virus, LIV = liver diseases, MHE = minimal hepatic encephalopathy, NASH = non-alcoholic steatohepatitis, OB = obesity, PAR = Parkinson's disease, PSA = psoriatic arthritis, RA = rheumatoid arthritis, T1D = type I diabetes, UC = ulcerative colitis. nonCDI controls are patients with diarrhea who tested negative for *C. difficile* infection. nonIBD controls are patients with gastrointestinal symptoms but no intestinal inflammation.
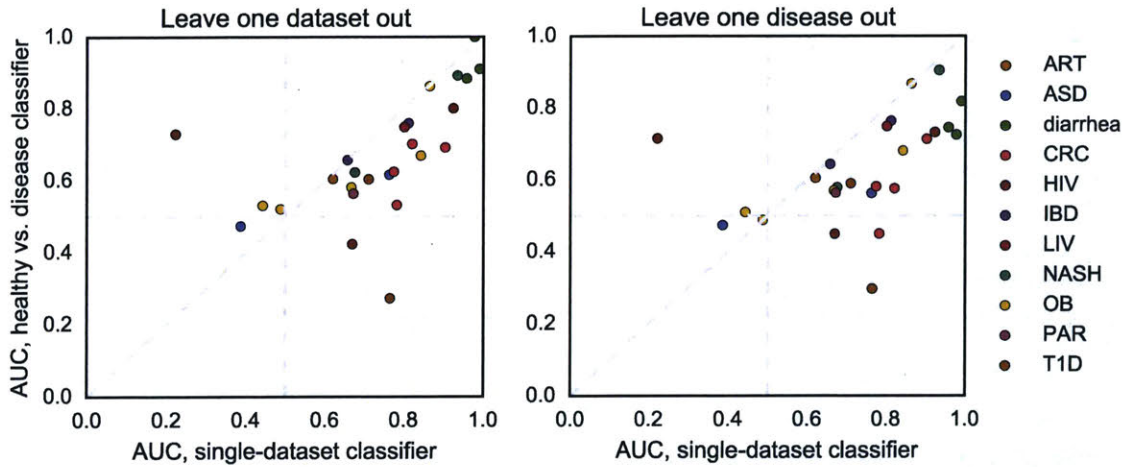
Figure B-9: *Both x-axes*: the area under the ROC curve (AUC) from each dataset's single classifier. Left: leave-one-dataset-out classifier. *y-axis*: the AUC of a classifier trained on all other datasets to distinguish healthy from unhealthy patients, tested on the left out dataset. Right: leave-one-disease-out classifier. *y-axis*: AUC from a classifier trained to distinguish healthy from unhealthy patients on all datasets except those of the tested disease. AUCs for each dataset were built from the classification probabilities on each test sample.

Figure B-10: The majority of non-specific microbes are robust to the exclusion of diarrhea datasets from consideration. The right-most bar shows order-level phylogeny, colored as in Figure 3A of the main paper. The left bar of the heatmap shows the original non-specific microbes, including all datasets. The middle bar shows the re-defined non-specific responders after excluding all diarrhea datasets. The right bar of the heatmap shows the non-specific microbes defined using Stouffer'âĂŹs method, combining one-tailed q-values across datasets and weighting by the square root of sample size (Stouffer combined q < 0.05).

Figure B-11: Empirical null distribution of the number of non-specific responders (colored points, x-axis indicates directionality of response), overlayed with the actual observed number of non-specific responders (black diamonds) for different defining heuristics (axis titles, i.e. "3 diseases" means that a genus needed to be significant ($q < 0.05$, Kruskal-Wallis (KW) test, Benjamini-Hochberg FDR correction) in three diseases in the same direction to be considered a non-specific responder). Empirical one-tailed p-values are printed above each distribution.
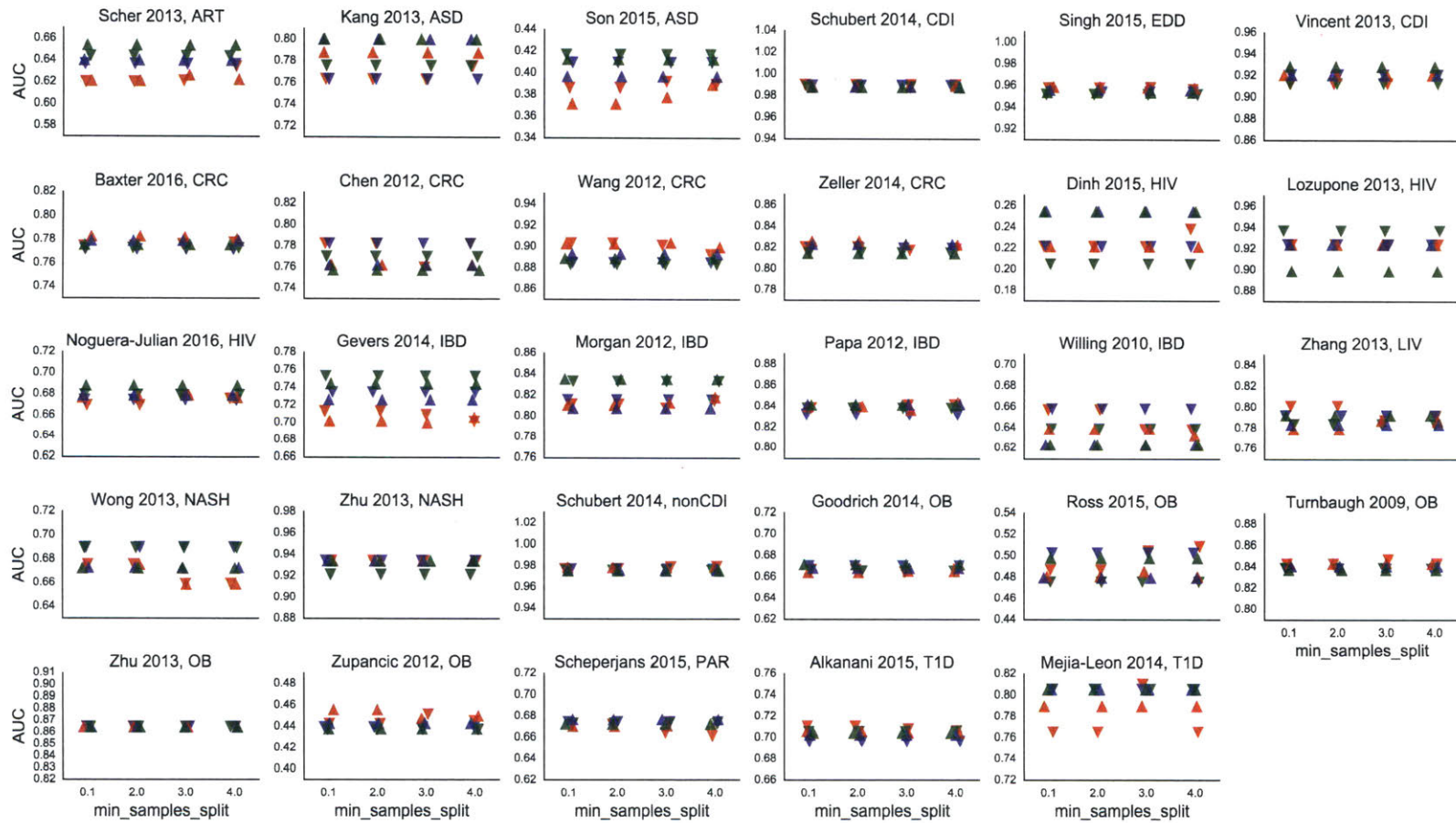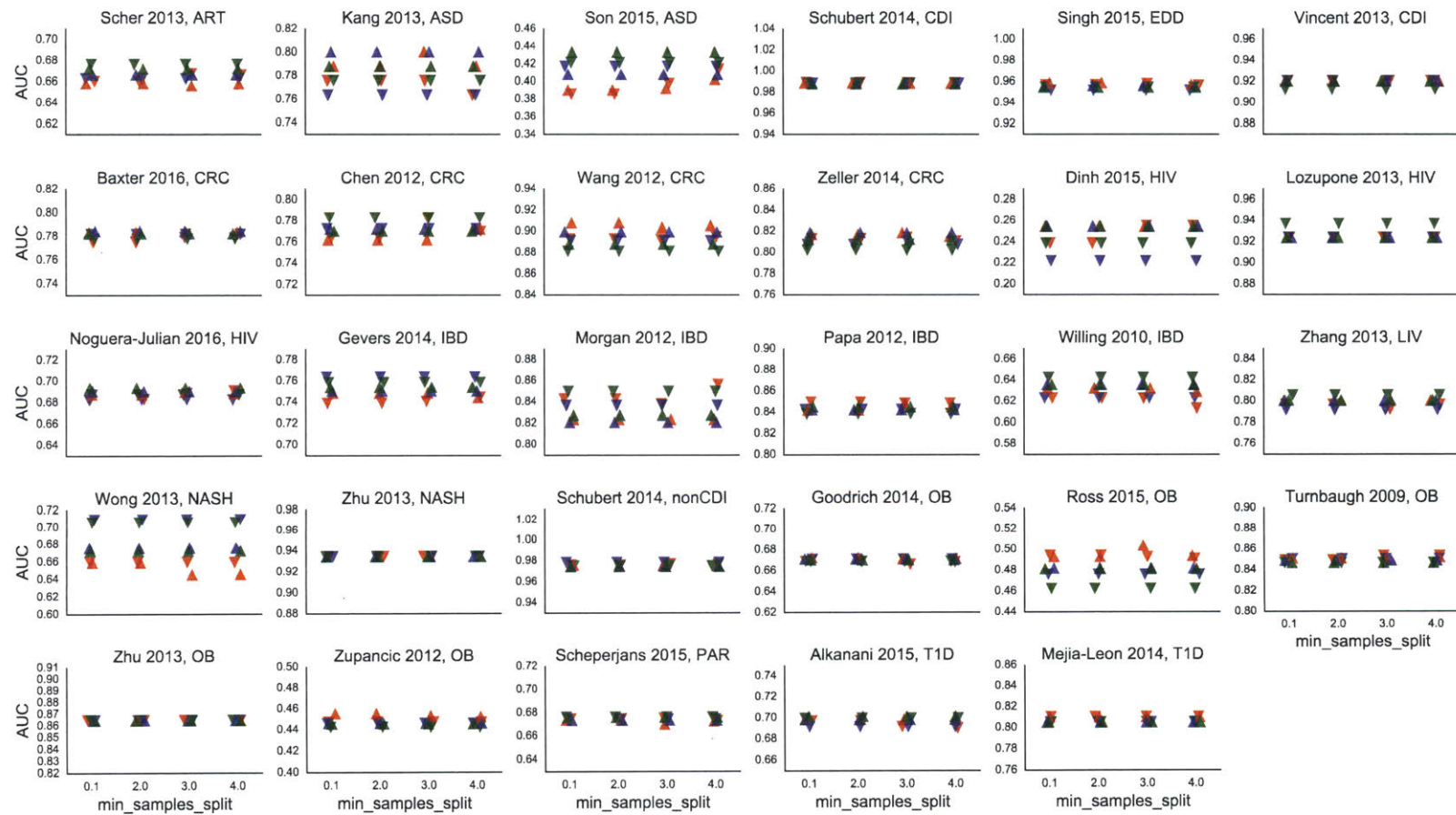
Figure B-12: Varying Random Forest parameters does not significantly affect area under the ROC curve in classifying cases from controls (Gini criteria). Random Forest classifiers built by using the Gini impurity ("gini") split criteria ("scikit-learn RandomForestClassifier"). Upward-pointing triangles are classifiers built with 10000 estimators; downward-pointing triangles are built with 1000 estimators. Colors indicate the value of `min_samples_leaf` (the minimum number of samples required to be at a leaf node): red = 1, blue = 2, green = 3. X-axes are the value of `min_samples_split` (the minimum number of samples required to split an internal node) [33]. All Random Forests were built using the random state seed 12345.

Figure B-13: Varying Random Forest parameters does not significantly affect area under the ROC curve in classifying cases from controls (entropy criteria). Random Forest classifiers built by using the entropy ("entropy") split criteria ("scikit-learn RandomForestClassifier"). Upward-pointing triangles are classifiers built with 10000 estimators; downward-pointing triangles are built with 1000 estimators. Colors indicate the value of `min_samples_leaf` (the minimum number of samples required to be at a leaf node): red = 1, blue = 2, green = 3. X-axes are the value of `min_samples_split` (the minimum number of samples required to split an internal node) [33]. All Random Forests were built using the random state seed 12345.
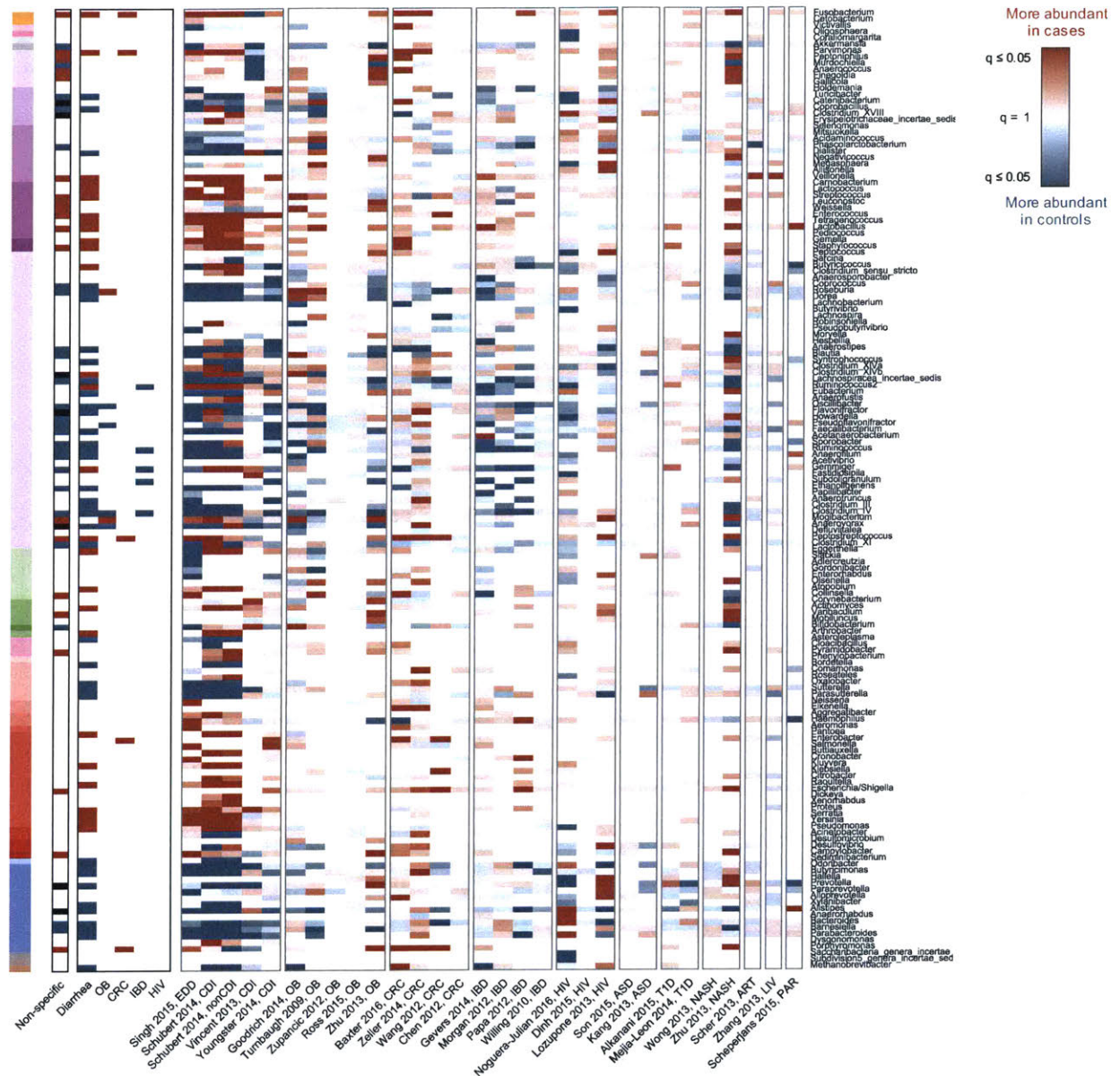
Figure B-14: Heatmap of log10(q values) for all genera which were significant (q < 0.05, Kruskal-Wallis (KW) test, Benjamini-Hochberg FDR correction) in at least one dataset, across all studies. Rows are genera, ordered phylogenetically (as in Figure 3A). Columns are datasets, grouped by disease and ordered according to total sample size (decreasing from left to right). The first and second heatmap panels from the left are the same as in Figure 3A. Q-values are colored according to directionality of the effect, where red indicates higher mean abundance in patients relative to controls and blue indicates higher mean abundance in controls. Opacity indicates significance and ranges from 0.05 to 1, where q-values less than 0.05 are the darkest colors and q-values close to 1 are gray. White indicates that the genus was not present in that dataset. ART = arthritis, ASD = autism spectrum disorder, CDI = *Clostridium difficile* infection, CRC = colorectal cancer, EDD = enteric diarrheal disease, HIV = human immunodeficient virus, IBD = inflammatory bowel disease, LIV = liver disease, NASH = non-alcoholic steatohepatitis, nonCDI = non-*Clostridium difficile* infection, OB = obesity, PAR = Parkinson's disease, T1D = type I diabetes.
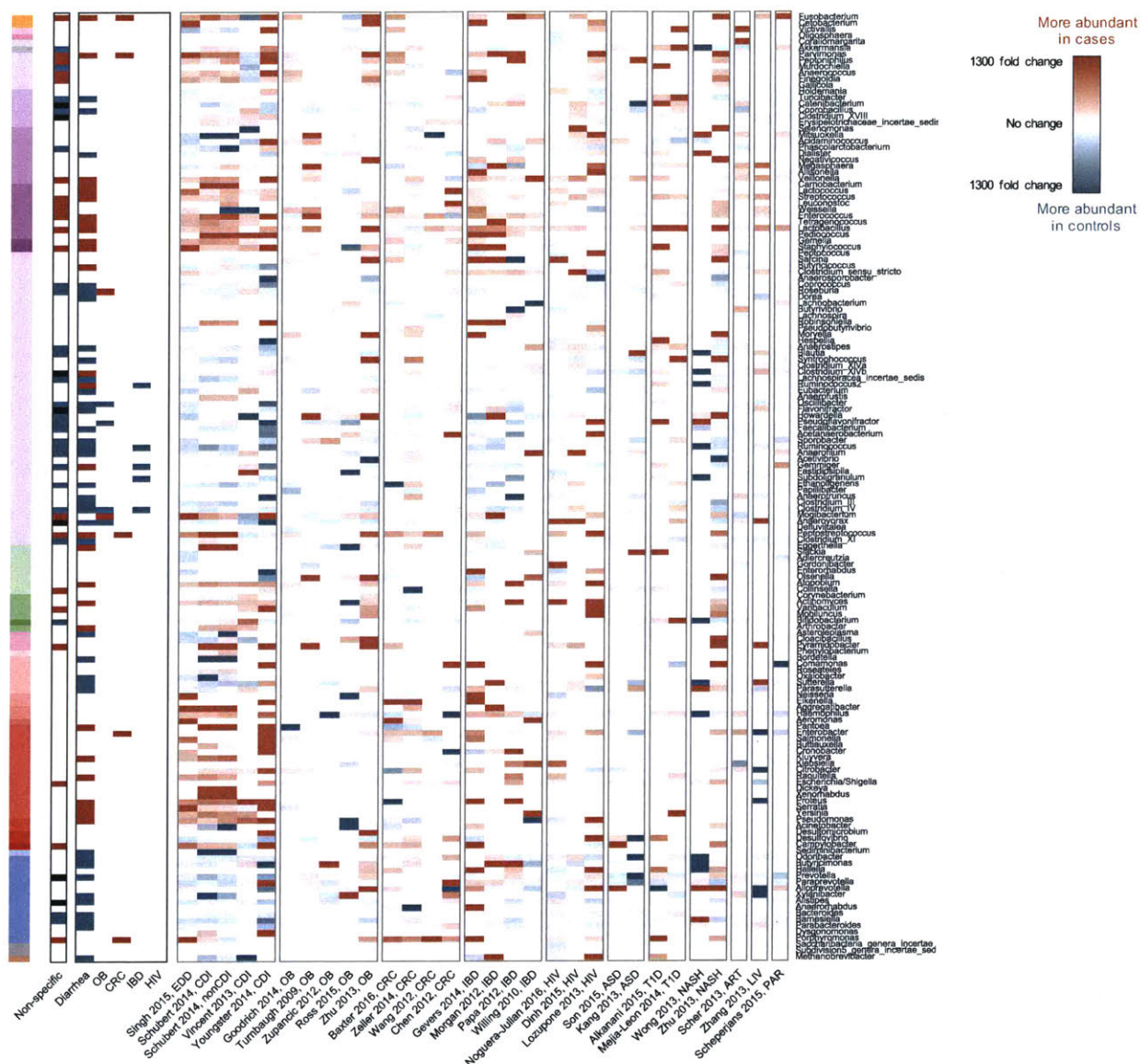
Figure B-15: Heatmap of log-fold change between cases and controls (i.e. $log_2(\frac{\text{mean abundance in cases}}{\text{mean abundance in controls}})$ for all genera which were significant ($q < 0.05$) in at least one dataset, across all studies. Rows are genera, ordered phylogenetically (as in Figure 3A). Columns are datasets, grouped by disease and ordered according to total sample size (decreasing from left to right). The first and second heatmap panels from the left are the same as in Figure 3A. Values are colored according to directionality of the effect, where red indicates higher mean abundance in patients relative to controls and blue indicates higher mean abundance in controls. Opacity indicates fold change and ranges from 1300 to 0, where fold changes greater than 1300 are the darkest colors and fold changes close to 0 are gray. White indicates that the genus was not present in that dataset. ART = arthritis, ASD = autism spectrum disorder, CDI = *Clostridium difficile* infection, CRC = colorectal cancer, EDD = enteric diarrheal disease, HIV = human immunodeficient virus, IBD = inflammatory bowel disease, LIV = liver disease, NASH = non-alcoholic steatohepatitis, nonCDI = non-*Clostridium difficile* infection, OB = obesity, PAR = Parkinson's disease, T1D = type I diabetes.

226

Bibliography

[1] A.M. Schubert, M.A. Rogers, C. Ring, J. Mogle, J.P. Petrosino, V.B. Young, D.M. Aronoff, and P.D. Schloss. Microbiome data distinguish patients with clostridium difficile infection and non-c. difficile-associated diarrhea from healthy controls. *mBio*, 5(3):e01021–14–e01021–14, 2014.

[2] C. Vincent, D.A. Stephens, V.G. Loo, T.J. Edens, M.A. Behr, K. Dewar, and A.R. Manges. Reductions in intestinal clostridiales precede the development of nosocomial clostridium difficile infection. *Microbiome*, 1(1):18, 2013.

[3] I. Youngster, J. Sauk, C. Pindar, R.G. Wilson, J.L. Kaplan, M.B. Smith, E.J. Alm, D. Gevers, G.H. Russell, and E.L. Hohmann. Fecal microbiota transplant for relapsing clostridium difficile infection using a frozen inoculum from unrelated donors: A randomized, open-label, controlled pilot study. *Clinical Infectious Diseases*, 58(11):1515–1522, 2014.

[4] P. Singh, T.K. Teal, T.L. Marsh, J.M. Tiedje, R. Mosci, K. Jernigan, A. Zell, D.W. Newton, H. Salimnia, P. Lephart, D. Sundin, W. Khalife, R.A. Britton, J.T. Rudrik, and S.D. Manning. Intestinal microbial communities associated with acute enteric infections and disease recovery. *Microbiome*, 3(1), sep 2015.

[5] Gregory P Donaldson, S Melanie Lee, and Sarkis K Mazmanian. Gut biogeography of the bacterial microbiota. *Nature Reviews Microbiology*, 14(1):20–32, 2016.

[6] N.T. Baxter, M.T. Ruffin, M.A. Rogers, and P.D. Schloss. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Medicine*, 8(1), 2016.

[7] Caitlin A Brennan and Wendy S Garrett. The gut microbiome, inflammation, and colorectal cancer. *Annual Review of Microbiology*, 70(1), 2016.

[8] G. Zeller, J. Tap, A.Y. Voigt, S. Sunagawa, J.R. Kultima, P.I. Costea, A. Amiot, J. Bohm, F. Brunetti, N. Habermann, R. Hercog, M. Koch, A. Luciani, D.R. Mende, M.A. Schneider, P. Schrotz-King, C. Tournigand, J.T. Nhieu, T. Yamada, J. Zimmermann, V. Benes, M. Kloor, C.M. Ulrich, M. von Knebel Doeberitz, I. Sobhani, and P. Bork. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Molecular Systems Biology*, 10(11):766–766, 2014.

[9] T. Wang, G. Cai, Y. Qiu, N. Fei, M. Zhang, X. Pang, W. Jia, S. Cai, and L. Zhao. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *The ISME Journal*, 6(2):320–329, 2011.

[10] W. Chen, F. Liu, Z. Ling, X. Tong, and C. Xiang. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS ONE*, 7(6):e39743, 2012.

[11] D. Gevers, S. Kugathasan, L.A. Denson, Y. Vázquez-Baeza, W. Van Treuren, B. Ren, E. Schwager, D. Knights, S. Song, M. Yassour, X.C. Morgan, A.D. Kostic, C. Luo, A. González, D. McDonald, Y. Haberman, T. Walters, S. Baker, J. Rosh, M. Stephens, M. Heyman, J. Markowitz, R. Baldassano, A. Griffiths, F. Sylvester, D. Mack, S. Kim, W. Crandall, J. Hyams, C. Huttenhower, R. Knight, and R. Xavier. The treatment-naive microbiome in new-onset crohn's disease. *Cell Host & Microbe*, 15(3):382–392, mar 2014.

[12] X.C. Morgan, T.L. Tickle, H. Sokol, D. Gevers, K.L. Devaney, D.V. Ward, J.A. Reyes, S.A. Shah, N. LeLeiko, S.B. Snapper, A. Bousvaros, J. Korzenik, B.E. Sands, R.J. Xavier, and C. Huttenhower. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol*, 13(9):R79, 2012.

[13] E. Papa, M. Docktor, C. Smillie, S. Weber, S.P. Preheim, D. Gevers, G. Giannoukos, D. Ciulla, D. Tabbaa, J. Ingram, D.B. Schauer, D.V. Ward, J.R. Korzenik, R.J. Xavier, A. Bousvaros, and E.J. Alm. Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease. *PLoS ONE*, 7(6):e39242, 2012.

[14] B.P. Willing, J. Dicksved, J. Halfvarson, A.F. Andersson, M. Lucio, Z. Zheng, G. J"arnerot, C. Tysk, J.K. Jansson, and L. Engstrand. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology*, 139(6):1844–1854.e1, dec 2010.

[15] J.K. Goodrich, J.L. Waters, A.C. Poole, J.L. Sutter, O. Koren, R. Blekhman, M. Beaumont, W. Van Treuren, R. Knight, J.T. Bell, T.D. Spector, A.G. Clark, and R.E. Ley. Human genetics shape the gut microbiome. *Cell*, 159(4):789–799, nov 2014.

[16] M.L. Zupancic, B.L. Cantarel, Z. Liu, E.F. Drabek, K.A. Ryan, S. Cirimotich, C. Jones, R. Knight, W.A. Walters, D. Knights, E.F. Mongodin, R.B. Horenstein, B.D. Mitchell, N. Steinle, S. Snitker, A.R. Shuldiner, and C.M. Fraser. Analysis of the gut microbiota in the old order Amish and its relation to the metabolic syndrome. *PloS one*, 7(8):e43052, 2012.

[17] P.J. Turnbaugh, M. Hamady, T. Yatsunenko, B.L. Cantarel, A. Duncan, R.E. Ley, M.L. Sogin, W.J. Jones, B.A. Roe, J.P. Affourtit, M. Egholm, B. Henrissat, A.C. Heath, R. Knight, and J.I. Gordon. A core gut microbiome in obese and lean twins. *Nature*, 457(7228):480–484, 2008.

[18] M.C. Ross, D.M. Muzny, J.B. McCormick, R.A. Gibbs, S.P. Fisher-Hoch, and J.F. Petrosino. 16s gut community of the cameron county hispanic cohort. *Microbiome*, 3(1):7, 2015.

[19] L. Zhu, S.S. Baker, C. Gill, W. Liu, R. Alkhouri, R.D. Baker, and S.R. Gill. Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH) patients: a connection between endogenous alcohol and NASH. *Hepatology*, 57(2):601–609, 2013.

[20] Marc A. Sze and Patrick D. Schloss. Looking for a signal in the noise: Revisiting obesity and the microbiome. *mBio*, 7(4), 2016.

[21] D.M. Dinh, G.E. Volpe, C. Duffalo, S. Bhalchandra, A.K. Tai, A.V. Kane, C.A. Wanke, and H.D. Ward. Intestinal microbiota, microbial translocation, and systemic inflammation in chronic HIV infection. *Journal of Infectious Diseases*, 211(1):19–27, 2014.

[22] Catherine A Lozupone, Marcella Li, Thomas B Campbell, Sonia C Flores, Derek Linderman, Matthew J Gebert, Rob Knight, Andrew P Fontenot, and Brent E Palmer. Alterations in the gut microbiota associated with hiv-1 infection. *Cell host & microbe*, 14(3):329–339, 2013.

[23] D.W. Kang, J.G. Park, Z.E. Ilhan, G. Wallstrom, J. LaBaer, J.B. Adams, and R. Krajmalnik-Brown. Reduced incidence of Prevotella and other fermenters in intestinal microflora of autistic children. *PloS one*, 8(7):e68322, 2013.

[24] J. Son, L.J. Zheng, L.M. Rowehl, X. Tian, Y. Zhang, W. Zhu, L. Litcher-Kelly, K.D. Gadow, G. Gathungu, C.E. Robertson, D. Ir, D.N. Frank, and E. Li. Comparison of fecal microbiota in children with autism spectrum disorders and neurotypical siblings in the simons simplex collection. *PLOS ONE*, 10(10):e0137725, 2015.

[25] A.K. Alkanani, N. Hara, P.A. Gottlieb, D. Ir, C.E. Robertson, B.D. Wagner, D.N. Frank, and D. Zipris. Alterations in intestinal microbiota correlate with susceptibility to type 1 diabetes. *Diabetes*, 64(10):3510–3520, 2015.

[26] M.E. Mejía-León, J.F. Petrosino, N.J. Ajami, M.G. Domínguez-Bello, and A.M.C. de la Barca. Fecal microbiota imbalance in mexican children with type 1 diabetes. *Sci. Rep.*, 4, 2014.

[27] V.W. Wong, C. Tse, T.T. Lam, G.L. Wong, A.M. Chim, W.C. Chu, D.K. Yeung, P.T. Law, H. Kwan, J. Yu, J.J. Sung, and H.L. Chan. Molecular characterization of the fecal microbiota in patients with nonalcoholic steatohepatitis – a longitudinal study. *PLoS ONE*, 8(4):e62885, apr 2013.

[28] Z. Zhang, H. Zhai, J. Geng, R. Yu, H. Ren, H. Fan, and P. Shi. Large-scale survey of gut microbiota associated with MHE via 16s rRNA-based pyrosequencing. *Am J Gastroenterol*, 108(10):1601–1611, jul 2013.

[29] J.U. Scher, A. Sczesnak, R.S. Longman, N. Segata, C. Ubeda, C. Bielski, T. Rostron, V. Cerundolo, E.G. Pamer, S.B. Abramson, C. Huttenhower, and D.R. Littman. Expansion of intestinal prevotella copri correlates with enhanced susceptibility to arthritis. *eLife*, 2, 2013.

[30] F. Scheperjans, V. Aho, P.A.B. Pereira, K. Koskinen, L. Paulin, E. Pekkonen, E. Haapaniemi, S. Kaakkola, J. Eerola-Rautio, P. Pohja, E. Kinnunen, K. Murros, and P. Auvinen. Gut microbiota are related to parkinson's disease and clinical phenotype. *Movement Disorders*, 30(3):350–358, dec 2014.

[31] Samuel Andrew Stouffer, Edward A Suchman, Leland C De Vinney, Shirley A Star, and Robin M Williams. Studies in social psychology in world war II. vol. I: The american soldier: Adjustment during army life. 1951.

[32] Marc Noguera-Julian, Muntsa Rocafort, Yolanda Guillén, Javier Rivera, Maria Casadellà, Piotr Nowak, Falk Hildebrand, Georg Zeller, Mariona Parera, Rocío Bellido, et al. Gut microbiota linked to sexual preference and hiv infection. *EBioMedicine*, 5:135–146, 2016.

[33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Appendix C

Supplementary Information for Chapter 6
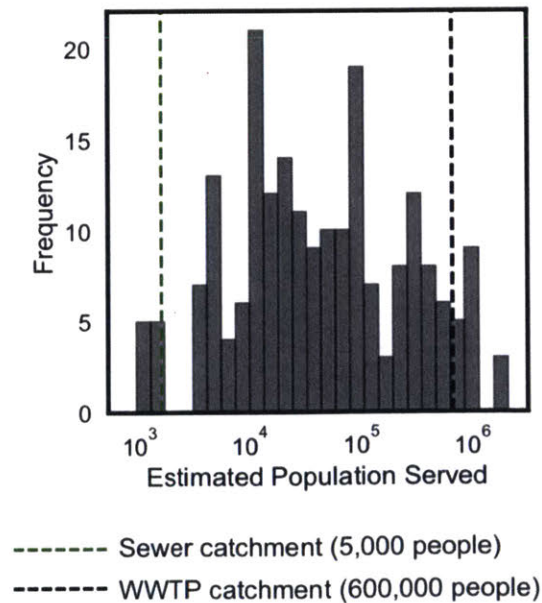
C.1 Supplementary Figures

Figure C-1: Distribution of catchment sizes in Newton et al, 2015. Green dashed line indicates catchment size of upstream residential manhole and black dashed line indicates estimated catchment size of downstream site sampled in this study.
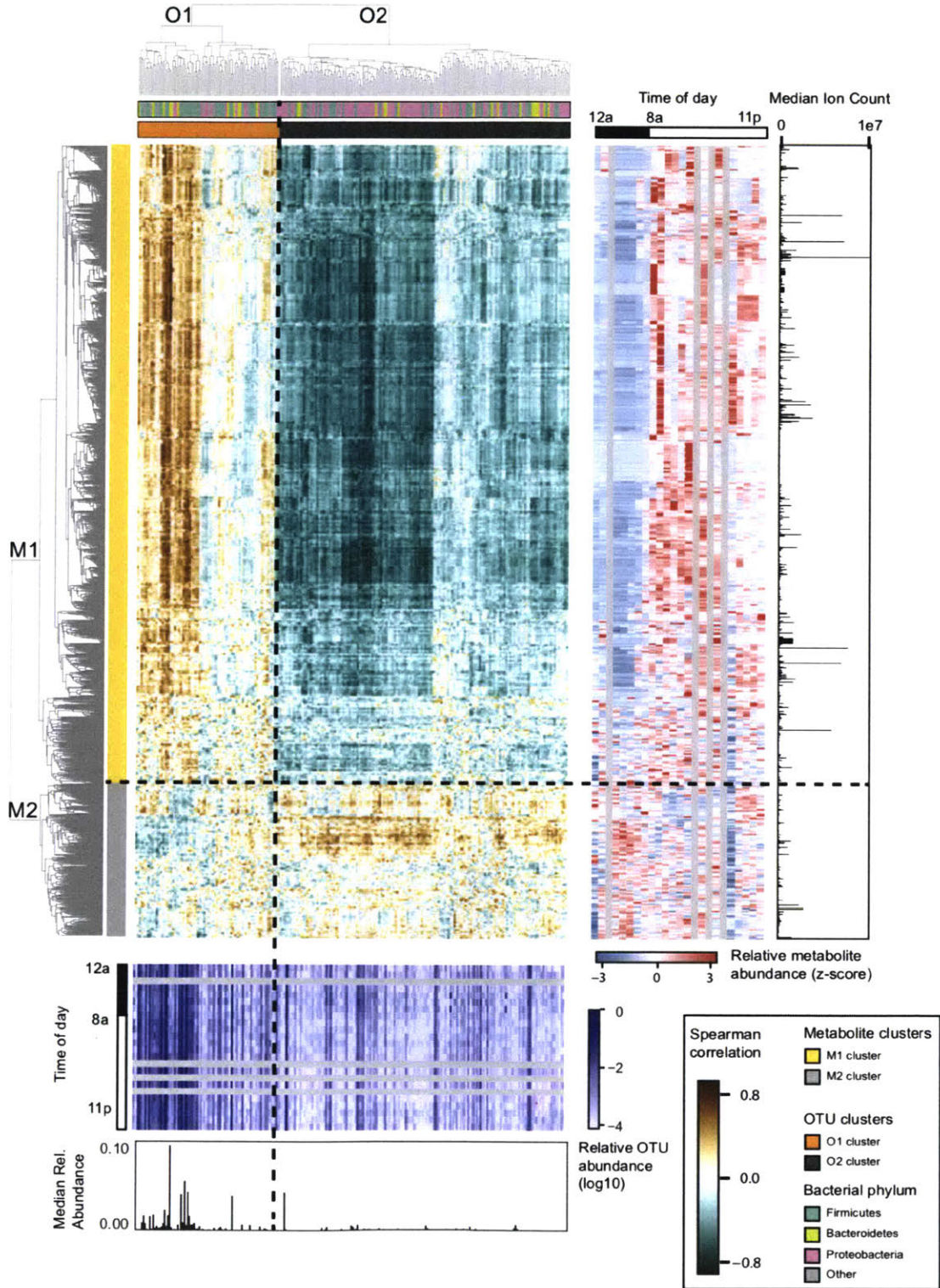
Figure C-2: Co-clustering metabolomics and microbiome data identifies human-associated and environmental clusters. (Middle panel) Spearman correlation between metabolites and OTUs. (Right panel) Z-scored metabolite abundances over 24-hour sampling. Metabolites are in rows, time points are columns. Right-most panel shows daily median ion count of each metabolite feature. (Bottom panel) Heatmap of log10(relative abundance) for each OTU. OTUs are in columns, time points are rows. Bottom-most panel shows daily median relative abundance of each OTU.
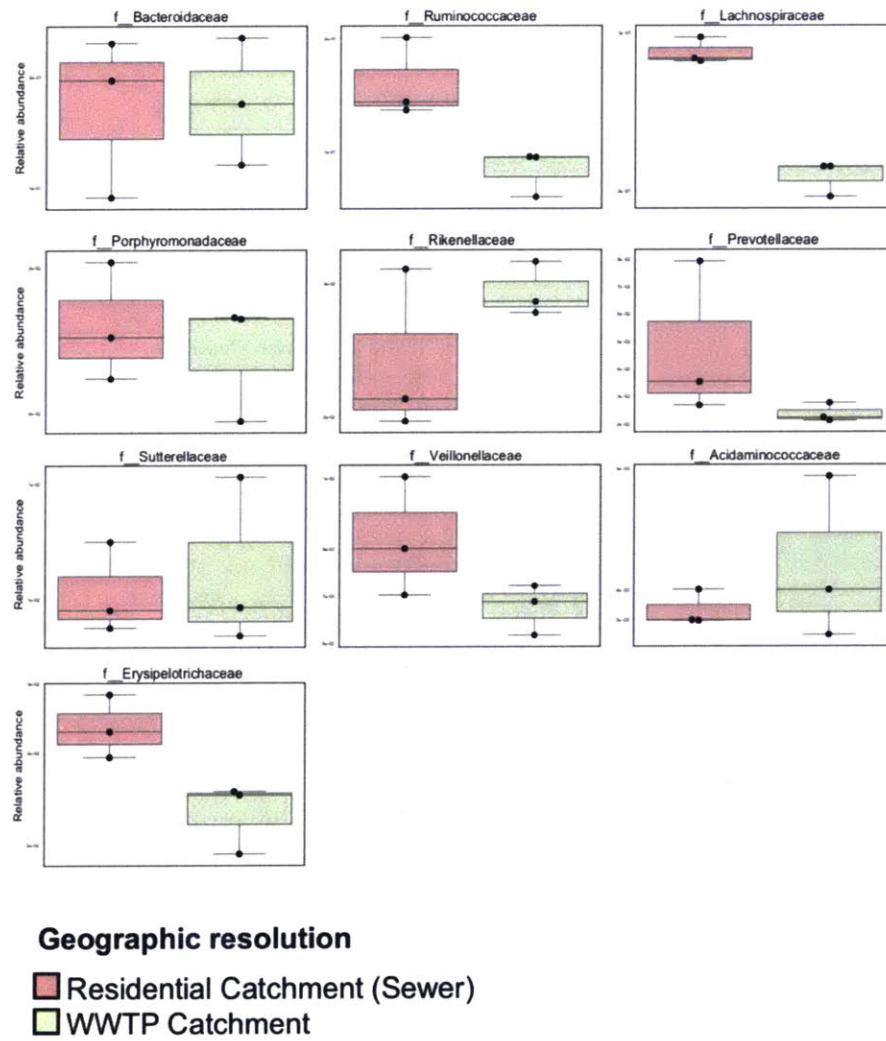
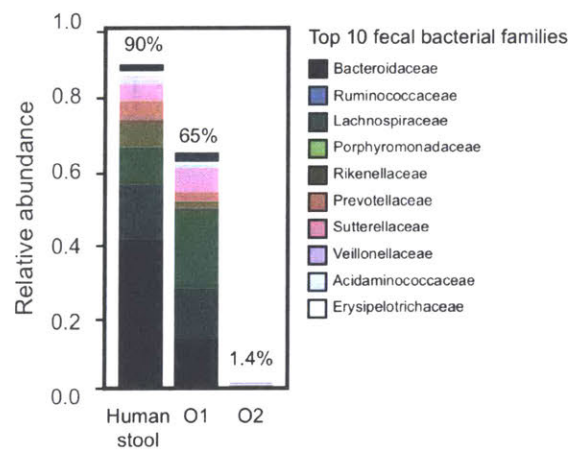Figure C-3: Abundance of human-associated bacterial families in upstream vs. downstream sample.

Figure C-4: Abundance of human-associated bacterial families in OTU clusters O1 and O2.