Improved Methods for Understanding the Gut Microbiome

Justine Woodhouse Debelius
University of Colorado Boulder, j.debelius@gmail.com

Follow this and additional works at: https://scholar.colorado.edu/chem_gradetds

Part of the Biochemistry Commons, Microbial Physiology Commons, and the Systems and Integrative Physiology Commons

Recommended Citation
https://scholar.colorado.edu/chem_gradetds/170
Improved Methods for Understanding the Human Gut Microbiome

by

Justine W. Debelius

B.S., Saint Mary’s College, Notre Dame, 2010
M.S., South Dakota State University, 2013

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Chemistry and Biochemistry

2015
This thesis entitled:  
Improved Methods for Understanding the Human Gut Microbiome  
written by Justine W. Debelius  
has been approved for the Department of Chemistry and Biochemistry

__________________________
Dr. Rob Knight

__________________________
Dr. Xuedong Liu

Date ________________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

IRB protocol #CU 12-0852; UCSD 141853
Debelius, Justine W. (Ph.D., Chemistry and Biochemistry)

Improved Methods for Understanding the Human Gut Microbiome

Thesis directed by Dr. Rob Knight

The gut microbiome plays a profound role in human health. Over the last decade, much work has been done to define differences between microbial communities in healthy and diseased individuals. Studies of healthy individuals have been small, and failed to address the multifactorial influences on the microbiome. However, little effort has focused on comparisons between the effects, in part due to the multivariate and complex nature of the data, which does not fit into the traditional paradigm for effect size calculations. This leaves questions about the relative impact of common practices on the microbiome, in comparison to the impact of a disease state. To address these challenges, I have combined novel and traditional approaches to microbiome analysis. I developed a method for estimating statistical power and effect size using Monte Carlo Simulation. The technique works as well as traditional approaches for parametric data, and out-performs the traditional methods when applied to nonparametric data. The observed effect sizes quantified previously observed biological conclusions. I applied the new power method to data from the American Gut Project, the largest open source, crowdfunded microbiome citizen-science project. The sample size within the American Gut made it possible to quantify and compare previously unobserved lifestyle effects on the microbiome. The effect of plant consumption on the microbiome was almost as large as the effect of antibiotic use in the last month, demonstrating the importance of diet in shaping microbial communities. The power technique was also applied to a study of Parkinson’s disease, to demonstrate the large effect associated with a Parkinson’s diagnosis. This work also represents the first time the influence of Parkinson’s disease status on the microbiome has been separated from the influence of Parkinson’s disease treatment. I propose a mechanism whereby the microbiome may be modulating dopamine production in Parkinson’s patients through a metabolite, butyrate. I then explore other examples of microbial metabolites modulating disease. Finally, I propose the use of
effect size calculation to identify targets for mechanistic investigation, followed by the application of multi-omics techniques to examine underlying pathways.
Dedication

To my grandparents

Mary Ann Hoffmann Woodhouse, Robert C. Woodhouse,

Betty Pass Debelius, Charles A. Debelius and Lillian McClaskey
Acknowledgements

This thesis would not have been possible without the help and support of many people. Thanks, first, to my advisor, Rob Knight, for opening the door to so many opportunities and continuing to support my research in areas in which I am passionate. Thank you to many lab mates who have mentored, guided, and provided invaluable advice on the techniques developed here. Special mention goes to Daniel McDonald, Chris Lauber, Embriette Hyde, Jessica Metcalf, Jamie Morton and Yoshiki Vazquez Baeza. Finally, nothing would be possible without the superheroes of the Knight lab: Greg Humpreys and the wet lab team, Gail Ackerman, Ulla Westermann, Jerry Kennedy, and Michiko Sousa.

Thanks to our amazing collaborators: Haydeh Payami and Erin Burns Hill on the Parkinson’s Project; Mark Asqueth from the Rosenbaum group; Tracy Gricksheit’s team, especially Kathy Schall; Sergio Baranzini, Patriccia Casaccia, Sarkis Mazmanian and Egle Cekanaviciute on the Multiple Sclerosis Project.

Thank you to my parents for their tireless support, even at ten o’clock at night when they had to get up at five am the next day. Love to Thomas and Harry, my wonderful ratties. Thanks to my siblings, and their encouragement to publish this document, and extend the theses on the feces of our species.

Thank you to my instructors, classmates and friends at CU Boulder. Thank you especially to Josh Shorenstein and Maria Balboa for reminding me both that there is life outside the lab, and that food is sometimes necessary for good science. Thank you also to Kim Kelley. This dissertation would not be possible without her constant, kind help.
## Contents

### Chapter

1. **Introduction**
   - 1.1 Critical considerations for obtaining crowdsourced data
     - 1.1.1 Participant Recruitment and Retention
     - 1.1.2 Sample Collection and Quality
     - 1.1.3 Metadata Collection and Quality
     - 1.1.4 Data Analysis and Dissemination
   - 1.2 Prospectus

2. **Monte Carlo Estimation of Statistical Power and Effect Sizes**
   - 2.1 The effect size problem
   - 2.2 Test based power and effect size
     - 2.2.1 Mathematical Definition of Empirical Power
     - 2.2.2 Pseudo Effect sizes
   - 2.3 Implementation
     - 2.3.1 Algorithm
     - 2.3.2 Metadata Pairing
     - 2.3.3 Assumptions
     - 2.3.4 Parameter Optimization
   - 2.4 Validation
2.4.1 Pseudo effect sizes .................................................. 25
2.4.2 Application to normal data and test based methods ..................... 28
2.4.3 Applications to non-normal data .................................... 30
2.4.4 Power extrapolation with underpowered datasets ........................ 30
2.5 Application to Microbiome Studies ..................................... 30
2.6 Prospectus ................................................................. 35

3 American Gut: an Open Platform for Citizen-Science Microbiome Research 38

3.1 Materials and Methods .................................................. 41
  3.1.1 Participant Recruitment and Sample Processing .......................... 41
  3.1.2 Primary Processing ................................................ 41
  3.1.3 Bloom Filtering .................................................. 43
  3.1.4 Statistical Analysis ............................................. 43
3.2 The American gut is diverse, and in some cases extreme ..................... 48
3.3 The Microbiome is influenced by lifestyle decisions and diet .................. 49
3.4 Lifestyle factors have different impacts on the microbiome .................. 59
3.5 Using the American Gut as a discovery platform ............................. 60
3.6 American Gut as a community process .................................... 63
3.7 Prospectus ................................................................. 65

4 Presence of Parkinson’s disease and therapeutic modalities have independent signatures in
the gut microbiome .......................................................... 66

4.1 Materials and Methods .................................................. 67
  4.1.1 Patient recruitment and metadata collection ............................ 67
  4.1.2 Microbiome Analysis ............................................. 68
4.2 Parkinson’s disease alters the microbiome community structure ............. 69
4.3 The effect of Parkinson’s disease is a combination of the disease itself and treatment 71
# Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Validation of pseudo effect sizes</td>
<td>26</td>
</tr>
<tr>
<td>2.2 Pseudo effect size from empirical power</td>
<td>29</td>
</tr>
<tr>
<td>2.3 Human study effect sizes</td>
<td>33</td>
</tr>
<tr>
<td>3.1 AGP sample type summary</td>
<td>39</td>
</tr>
<tr>
<td>3.2 AGP demographics</td>
<td>40</td>
</tr>
<tr>
<td>3.3 Sequences used in bloom filtering</td>
<td>45</td>
</tr>
<tr>
<td>3.4 AGP Alpha Diversity Regression</td>
<td>58</td>
</tr>
<tr>
<td>3.5 AGP paired effect sizes</td>
<td>62</td>
</tr>
<tr>
<td>4.1 Taxa associated with Parkinson’s disease and age</td>
<td>71</td>
</tr>
<tr>
<td>4.2 MRM of treatment in Parkinson’s patients</td>
<td>72</td>
</tr>
<tr>
<td>4.3 Parkinson’s diagnosis and treatment alter microbial community structure</td>
<td>73</td>
</tr>
</tbody>
</table>
Figures

Figure

2.1 Monte Carlo power algorithm .................................. 18
2.2 Model for subsampling based on metadata control categories ........ 20
2.3 Power collapses as a function of the number of tests and number of power calculations 22
2.4 Combinatorial depth as a function of sample and subsample size ........ 25
2.5 Pseudo effect sizes ........................................... 27
2.6 Monte Carlo estimation of parametric power .......................... 29
2.7 Monte Carlo estimation of nonparametric power ..................... 31
2.8 Subsampled data recapitulate effect sizes .......................... 32
2.9 Effect sizes in human microbiome studies .......................... 34
2.10 Study effects are larger than biological effects ..................... 36
3.1 Participant results certificate .................................... 42
3.2 Bloom filtering alter the relative abundance of Gammaproteobacteria .......... 44
3.3 Diversity in the AGP ........................................... 50
3.4 Lifestyle influences in the American Gut Project .................... 51
3.5 Sex effects on alpha diversity in American Gut ..................... 52
3.6 Alpha diversity comparisons in the American Gut .................... 54
3.7 Beta diversity comparisons in the American Gut ..................... 55
3.8 American Gut co-occurrence network ............................ 56
3.9 Significant associations between KEGG orthologies and metadata ..................................... 57
3.10 American Gut power curves .................................................................................................. 61
3.11 American Gut compared to ICU samples ................................................................................. 64
4.1 Parkinson’s Disease has a large effect on the microbiome. ................................................. 70
5.1 The production-absorption-target site pathway of metabolites ......................................... 79
5.2 Specialized metabolites, synthesized by the gut microbial community from various
    precursors, reach target sites and mediate health and disease ............................................... 86
6.1 Transgenic rats rats separate from Wild Type in Unweighted UniFrac PCoA Space. .............. 90
6.2 The time of rearing drives differences in PCoA space between Fischer Rats ....................... 91
6.3 The Cause of SBS changes the microbiome between individuals ........................................... 94
6.4 Treatment status alters the microbiome in MS patients. .................................................... 96
Chapter 1

Introduction


The human microbiome, or the collection of microorganisms in, on, and around humans, are poorly understood, but as recent evidence has shown, a critical component of health. We host more than 90% of the cells in our bodies, and microbial symbionts contribute 100 times more genes than our human genome (1; 2; 3).

Modern microbiology began with Van Leeuwenhoek and Robert Hooke’s observation of microorganisms (4). The nineteenth century “Golden Age” of microbiology had its roots in the work of luminaries including Robert Koch, Louis Pasteur, Edwin Klebs, John Lister, Walter and Fanny Hesse, and many others (5; 6; 7; 8). Their work focused on the microorganisms which caused disease using culture-based studies and observation (5; 6). However, culture based techniques are unable to capture the full scale of microbial diversity (8; 9; 10).

The target for modern culture-independent microbiome analysis comes from a short 1977 communication by Woese and Fox (11), which identified the the small Ribosomal RNA (rRNA) as a universal marker for phylogeny. This paper also identified three separate domains of life: Archaea, Bacteria and Eukaryota. Work through the 1980’s and 1990’s expanded understanding of disparate domains of life and culture independent approaches (reviewed in (9)). Early sequencing efforts were limited by the technology, and culture independent methods benefited greatly from advances in next generation sequencing. The 2004 commercial release of the Life Science 454 platform made
faster, deeper sequencing possible for many fields, including the microbiome (12).

In June 2005, Eckburg and colleagues (13) published the first culture free study of the human gut microbiome, identifying 395 organisms, 80% of which were previously unseen. This was shortly followed by the first observation that obesity altered the gut microbial communities of mice (14), and confirmed in people the next year (15). In the intervening decade, the microbiome has risen in prominence in the medical world. Influences on microbiome community structure, or the type of organisms present and their relative abundances, is influenced by many factors, including host genetics, diet, and xenobiotic and antibiotic use (16; 17; 18; 19). The gut microbiome, in particular, plays an important role in metabolism, immune system development, and endocrine and neurological signaling (20; 21). Dysbiotic gut communities have been associated with a host of human diseases including obesity, inflammatory bowel disease, type I and type II diabetes, autism, multiple sclerosis, and malnutrition (22; 21).

Predictive models using gut microbiome data can assess risk for conditions like Crohn’s disease (23) and obesity (24). Furthermore, microbial transplants may also transmit clinical phenotypes in some cases: one report suggested a donor transmitted a risk for obesity to her human recipient along with her fecal microbiome, while trans-species transmission of obesity and malnutrition is well established (25; 19; 26).

Human microbiome work has primarily focused on case-control studies of a few dozen to a few hundred individuals. Budget restrictions and a strict disease focus by funding agencies often limit the size and scope of investigations. Although studies supported by traditional mechanisms have led to considerable advances, there are also major pitfalls associated with the traditional approach. Small cohorts create inconsistent observations among studies. Trends in community structure are often shared between studies, but the individual taxa driving these trends often are not. For example, studies have found correlations between obesity and both an increase and a decrease in *Methanobrevibacter smithii* (27). Meta-analyses can ameliorate inconsistencies due to different approaches in data analysis, but cannot correct for differences due to sample handling or the characteristics of the control and clinical groups (28). The problem is compounded by
the absence of effective, mathematically justified ways to quantify effect size or the signal-to-noise landscape in the microbiome.

Even previous efforts to establish a “healthy” microbiome baseline have been small compared to the cohorts used in other correlative analyses, like Genome Wide Association Studies. The Human Microbiome Project (HMP), a cohort commonly used as a healthy reference, focused on 242 healthy professional students in their twenties and thirties living in two regions of the United States (29). The HMP contributed valuable information about the microbiome, including the variation in taxonomic abundance in healthy adults and the lack of a core healthy microbiome. However, the HMP did not answer all the open questions about the healthy microbiome, and indeed posed more questions. In comparison with a study comparing children and adults in the US, Malawi, and Venezuela, HMP participants captured only a small subset of the total diversity in that study (30).

Public participation in microbiome research, through crowdsourcing and crowdfunding, may provide some potential solutions to these problems. Both participatory models open the project to the public, making science accessible to a wide group, rather than a practice for experts in semi-isolation. Crowdfunding enables public participation science by asking for a monetary investment in a project. Lay people can determine what they consider worthy or unworthy of funding, whether it be comparative studies of the cat microbiome (31) or a qualitative survey of the best burritos in San Francisco (32). Meanwhile, crowdsourcing engages the public in data collection or analysis. Individuals participate by contributing data, physical samples, or analyses. This typically involves the contribution of observational data, such as flu symptoms or bird sightings, in projects such as Flu Near You (33) or the annual Audobon Christmas Bird Count (34), but may also involve crowdsourced data collection, modeled by the Personal Genome Project (35), or even crowdsourced data analysis, through platforms like the online games, Foldit (36; 37) and EteRNA (38; 39). Crowdsourcing may present opportunities to access populations, areas, or information difficult for a finite group of researchers to access. It also offers opportunities in exploratory science, since the wealth of data allows for a degree of exploration that is more difficult to achieve in traditionally sourced studies where participant recruitment is often more focused.
The American Gut Project (AGP) (40) is a crowdfunded, crowdsourced microbiome project. The AGP began at the Biofrontiers Institute at the University of Colorado, Boulder, and is now based at the University of California at San Diego. It was initiated as a collaboration between the Earth Microbiome Project (EMP) (41) and the Human Food Project (42). Participants provide a monetary contribution to help cover the costs of the project, and are then sent a kit for physical sample collection (fecal, oral, skin, pet, or environmental). Following informed consent, participants answer a survey about their health, and send in their sample. Individual participants receive a report describing their results. De-identified data is also deposited into the public domain without usage restrictions. We have used American Gut data to draw conclusions about factors associated with human health, which are presented in Chapter 3. Here, we present three stages that have been important for aggregating the American Gut results and presenting usable data.

1.1 Critical considerations for obtaining crowdsourced data

Communication is central to successful science, especially crowdsourced science. There is an added complexity in disseminating research to the general public, because complex concepts must be translated into messages that can be readily digested by individuals without specific domain knowledge. The inherent difficulty is magnified in participatory science, as there is a continual interaction with members of the general public. The communication challenge can be broken down into three majors areas critical to crowdfunding: participant recruitment and retention, data collection, and data dissemination.

1.1.1 Participant Recruitment and Retention

The first area, recruitment and retention, is a crowdsourced project’s initial and primary interaction. At the outset, members of the public are unlikely to be interested in your project if they are unable to understand why you are doing the project in the first place. They also want to know how they benefit by participating. In the case of the American Gut Project, one of our goals was to provide an avenue in which members of the general public could engage in cutting-edge
research, and in turn, learn about the organisms that inhabit their bodies.

Participatory science can be self-selecting, and this may create a biased cohort, rather than a true representation of a targeted population. Gut microbiome research, for example, may be more likely to attract individuals with diagnosed gastrointestinal conditions, such as inflammatory bowel disease. In the American Gut, we see a ten-fold enrichment in participants with IBD compared to the US population (see Chapter 3). Many of the early American Gut Participants were individuals who emphasized the importance of diet in health, and therefore tended toward more extreme dietary choices. Sponsors have also contributed funds to provide American Gut kits for participants in other populations of interest, including children with Autism Spectrum Disorder. These sub-studies may lead to an understanding of compositional patterns associated with these specific populations.

Other, less explicit biases, may also appear in the data. The role of the internet in participatory science cannot be discounted, meaning that participation is likely linked to internet access (43). Coupling crowdfunding to crowdsourcing may also limit the participant population to those able to afford the cost or may even lead to self-selection among those with the available disposable income. These implicit biases in the population may be hard to identify, and harder to correct (although they are less important for the original goal of the project in terms of identifying the diversity of types of microbiome “out there in the wild”). Decoupling crowdsourcing and crowdfunding, at least for some cohorts, may help ameliorate some biases in data.

A second important interaction with participants arises when, inevitably, participants have questions. Mismanagement of the participant base can be a major reason projects fail (44). The help burden stems not just from the number of questions coming in, but the number of personnel hours necessary to answer these questions. Given the nature of crowdfunding, it is not known in advance the scale to which a project will grow, which makes scoping personnel effort difficult and risky (i.e., if the project “fails”). In microbiome research, the potential for health discoveries adds a new level of complexity. Participants and backers may choose to engage a project in which the research personally benefits them. For the American Gut Project, despite all our efforts at dissuading participants of the notion that the data generated have current medical value, we still
frequently receive questions along the lines of “I have condition X, given my microbiome, what do you recommend I do?”

1.1.2 Sample Collection and Quality

Once participants are recruited, the next major hurdle is collecting their data. In microbiome studies, this typically involves a physical sample, or set of physical samples, and information about the participant and sample. Physical sample collection poses a challenge for biologically-based projects. The sampling protocol needs to be simple and safe. However, even simple protocols can be complicated for novices without clear instructions. The unfortunate reality is that people are bad at following instructions (for example, we anticipate that few of the readers of this article read their cell phone manual cover to cover). Explicit, succinct, and engaging instructions are vital to minimize variability in how the instructions are followed. To this end, the American Gut Project took two approaches. The first was an eye-catching “quick instructions” sheet that gave a rundown of the steps necessary for sample collection and return. In addition, detailed instructions were provided as well, including video examples on the web site. During the course of the project so far, it has been necessary to revise the instructions based on feedback from participants, and address obvious issues with sample collection. Notably, we discovered that the amount of fecal matter to send in was ambiguous, leading us to provide graphic examples of good and bad samples. As we refined the instructions, we encountered fewer questions and higher quality samples were returned.

1.1.3 Metadata Collection and Quality

The human microbiome is contextually dependent, making it impossible to understand a microbiome community without information about its host (45; 46). Therefore, participant and sample metadata (i.e. contextual information) are a second important consideration in participatory microbiome research. The goal of metadata collection is to maximize the amount of accurate, usable data that can be collected for every sample. Survey design and implementation can support or impede this end. Although it is possible to analyze a few dozen free response fields for a
small number of samples, it is prohibitive to analyze large numbers of free-response fields for large numbers of samples. Free response fields are also more likely to contain human error. For example, in the American Gut dataset, individuals have reported chicken as their most common carbohydrate, which would be surprising if true (standard nutritional data for chicken breast report zero carbohydrates). Questions with controlled vocabulary, such as multiple choice questions or fields limited to accept bounded numeric responses, can help improve accuracy. It may also be important to consider the level of detail that is possible to record in a survey. Controlled vocabulary represents one of these trade-offs. Another is the decision whether or not to pursue information about a specific medical condition. The American Gut has addressed these issues with triggered response questions, condition-specific surveys, and the option to follow up with participants.

Metadata errors are inevitable - whether in self-reported data or well-designed clinical studies (47). There are two major considerations with error reporting: how the errors are identified, and the way in which errors are corrected or removed. Identifying obvious errors can be easy. In the American Gut, participants who reported birth dates prior to the start of the twentieth century were identified as obvious errors. There are also profound differences between adult microbial communities based on body site, which can help when participants forget which sample was collected on which swab (29). However, other errors can be more difficult to identify. In certain American Gut analyses, we noticed that alcohol had a larger effect than antibiotic use, and that infants (0-3 years of age) had microbiomes that were more diverse than older children, a contrast with previous publications (30). When we examined the infant data further, we identified several individuals with age listed as less than three years of age, but self-reported height over four feet and reported drinking more than once a week, leading us to question the accuracy of the reported ages. In a large dataset, it can be useful to remove clearly erroneous information, especially if the correct answer is difficult to determine. Age values that are likely incorrect, given the rest of the contextual information, are therefore removed from analysis with the American Gut data. Mislabeled body sites can be corrected even against a high background mislabeling rate using a supervised learning technique, due to the strength of the association between body site and community structure (48).
The same associations may prove true for other parameters as we continue to collect data.

1.1.4 Data Analysis and Dissemination

Data dissemination and communication are the final steps in the scientific process. In a traditional scientific model, this has taken the form of publication in grant reports, scientific journals, and the deposition of data to repositories. Participatory science opens questions about data ownership, dissemination, and communication. Rather than delivering results to a grant committee of peers, scientists instead must communicate results to a wider community. However, in crowdsourced projects, individualized results may be offered as an incentive for participation. In human microbiome research, it may be challenging to balance providing novel results while avoiding presenting information that could be interpreted as a medical diagnosis. In crowd-funded projects, regular updates showing progress are important to continued investment and reinvestment (44); for a scientific project, this can mean everything from a blog with regular updates to a public release of data and analysis techniques.

Providing aggregated crowdsourced data to the general public can also crowdsourcethe analysis. It sends a clear message that the public owns the data. Large datasets present opportunities for exploration, new technique development, and technique refinement. Providing the dataset to a collaborator network early on fosters opportunities for new analyses and directions. Collaborations that play on the strengths and expertise of each group can accelerate the rate of discovery. Making the full dataset available through open access mechanisms early in the analysis process is one of the simplest ways to disseminate data to multiple collaborators at a variety of institutions.

However, data release can raise privacy concerns. Institutional Review Board (IRB) protocols must make clear how the participant’s de-identified data can and should be used. Participants’ de-identified microbial DNA sequence data and per-sample and per-individual metadata will be made publicly available, if that is a goal of the project. Releasing data into repositories without monitoring may make dissemination easier, but it can also mean that after participants withdraw, their data cannot be retracted. Additionally, extensive care has to be taken to avoid compromising
the anonymity of the participants. Such steps include separating clearly identifying participant
data from survey information; limiting access to raw data; and removing identifying information
from publicly available survey results, even inadvertently identifying information. To this end, the
surveyed data must be validated against possible identification threats, for example a combination
of date of birth and zip code, could provide an attacker with identified personal information of a
participant.

1.2 Prospectus

Crowdfunding and crowdsourcing, while powerful ways to fund projects, recruit participants,
and raise public awareness and interest, are novel approaches with their own pitfalls. The nature
of a crowdfunded projects requires different approaches from traditional study designs and consid-
erations, especially with respect to public relation and communication. Defining the intention and
standing of the project is vital when individuals have a personal and financial stake. Communi-
cation of the project expectations, what the participant can expect to receive, and progress of the
project and individual participant samples, especially if there is a waiting period between financial
contribution and tangible results, cannot be overlooked.

Barriers to participant and the associated participation bias must be considered. Crowd-
funded research projects are almost certainly expected to draw in a specific subset of the population,
leading to potentially biased sampling. The financial cost of participation may exclude an addi-
tional subset (although this can be ameliorated by supplementing crowdfunding by philanthropic
contributions and/or foundation support). Additionally, considerations of how to reduce and re-
spond to errors in the data must be considered. Data dissemination, in the form of individualized
results, and sharing analysis tasks, can also benefit or hinder projects.

In summary, citizen science provides a new opportunity for microbiome research. While
crowdfunding is unlikely to replace grant funding from government and private agencies, it may act
as an additional mechanism for answering questions that are difficult to explore through traditional
means.
Chapter 2

Monte Carlo Estimation of Statistical Power and Effect Sizes

In the last chapter, I discussed some of the challenges inherent in obtaining data from large cohort studies. However, such studies present tremendous opportunities, because by collecting many variables from the same subject population, we can start to understand which of these variables have large effects, which have small effects, and which interact to shape complex microbiomes. A standardized comparison between studies allows us to validate effect sizes across cohorts, independently of the effect size. Factors with large effect sizes may create better targets for future mechanistic work. However, inherent to the problem of calculating microbiome effect sizes is the issue of working with highly multivariate, compositional datasets. In this chapter, I introduce a new technique to address this issue. I developed, coded, validated and applied this method.

2.1 The effect size problem

Many current microbiome studies are small cohort studies, focused on particular variables of interest. Due to the sample sizes, results and conclusions may be contradictory, depending on the cohort in question. For example, it is unclear whether or not there is a change in alpha diversity, a measure of within-community diversity, associated with Type I Diabetes. A Finnish study comparing eight children with and eight children without diabetes over three time points found that diabetic children had different trajectories, leading to a significantly lower alpha diversity at the third time point (49). In contrast, a study of 29 Mexican children at a single time point found no significant difference between children who had been recently diagnosed with diabetes,
those who had been diabetic for more than two years, and age-matched healthy controls (50). It is difficult to determine if the differing conclusions are due to differences in the study design, cohorts, methodology, or simply small sample size.

In larger studies, it can be difficult to rank the strength of effects on the microbiome, unless they drive separation in principal coordinates (PCoA) space. However, PCoA projections are not always reliable for effect size comparisons. In the largest microbiome study of treatment naive Crohn’s patients \(n=668\), there was no clear separation in PCoA space between children with Crohn’s Disease and those without any inflammatory bowel disease diagnosis (IBD), although several taxa differentiated the microbiomes (23). As previously reported, antibiotic use exasperated dysbiosis, and could be weakly correlated with disease severity. Based on differences in p values between cohorts with vastly different sizes, the authors concluded there was a weak effect associated with antibiotic used - despite no formal estimation.

The adoption of effect size and statistical power within microbial ecology could help to standardize both within and between study comparisons. Within a study, effect size calculations can be used to rank predictive factors, which in turn, can inform the design of future observational and mechanistic studies. Between studies, effect size calculations can provide a basis of comparison, independent of the number of observations used, although effect sizes cannot correct for real differences between cohorts due to technique effects. Nakagawa and Cuthill (51) argue that effect sizes help move away from an artificially dichotomous rejection of a hypothesis based on an arbitrary critical value, better acknowledge noise within a dataset, and allow the gradation of biologically relevant effects. Unlike p-values alone, which depend on the sample size, effect sizes can easily be compared between studies and cohorts. Additionally, expanding a library of effect sizes can help shape future studies; power and effect size are closely tied (52). As the number of effect size calculations expands, it becomes easier to estimate appropriate sample size, leading to better study design overall.

The three most common approaches to microbiome characterization - taxonomic composition, alpha diversity and beta diversity - all present challenges in terms of power calculations. The
Composition is typically described in terms of Operational Taxonomic Units (OTUs), a cluster of marker gene sequences roughly representing genus level similarity (53; 54). OTUs are sparse, which is to say that an individual OTU may be absent in a large fraction of communities in a biome (55). Individual taxon abundances are typically modeled using poisson or negative binomial distributions (55; 56). Additionally, OTU abundance cannot be mapped back to the actual abundance of an organism due to variable copy number, primer and extraction bias, and nonquantitative PCR (53; 57; 58; 59; 60). The approach to normalization can affect the quality of the model used to evaluate taxonomic differences (61). Common approaches to this problem include subsampling to even depths and applying nonparametric tests (i.e. Mann Whitney U and Kruskal-Wallis) and differential abundance methods adapted from RNA sequencing (55; 61; 62). Methods papers do not typically define a power formula based on the underlying distribution for the test, as can be seen with parametric tests like t tests (52). Instead, power-based comparisons are evaluated using simulated communities (i.e. (61; 63; 64)). The simulations are a useful heuristic for comparison of the methods, but are not practical for power and effect size determination on real data.

Although there are many metrics to quantify alpha diversity, few methods exist to specifically compare those quantities. It is well known that alpha diversity depends on an even subsampling depth, so care must be taken to ensure an even depth across all communities (53). However, no consensus has been reached about the normality of the data, and so parametric, nonparametric and permutation tests are all commonly applied to alpha diversity comparisons. Power and effect sizes based on parametric tests may be most appropriate for alpha diversity, although they are still not commonly used in microbiome research.

Beta diversity is a commonly used concept in ecology that facilitates the comparison of communities across ecosystems. Common metrics used in microbiome studies, like unweighted UniFrac distance (65), weighted UniFrac distance (66), or Bray-Curtis distance (67), are expressed as a dissimilarity or distance matrix, providing pairwise distances between samples. The nature of these distances is such that individual distance observations are not independent, and therefore provide challenges to test. Anderson (68) proposed a permutational ANOVA (permanova). The
ANOVA test statistic does not follow a known distribution, making traditional, test-based methods challenging. A power method for permanova was proposed earlier this year, which relies on sampling the distance matrix with replacement to estimate power (69).

To address the challenges inherent in microbiome analysis, we have developed a Monte-Carlo estimation of statistical power using sampling without replacement. The method is test-agnostic, comparable to parametric, test-based methods for normal data, and out performs test-based methods for nonparametric data. The technique can readily be applied to microbiome studies, and may be particularly helpful to identify future targets for mechanistic work or technique refinement.

2.2 Test based power and effect size

Statistical power measures the ability to distinguish correct from incorrect outcomes of a statistical test, for a given critical value, sample size, and effect size. Traditionally, power has been defined using test-based methods, relying on knowledge of the test statistics and underlying distributions. (70). Let \( \{S_1, S_2, \ldots, S_k\} \) be a set of samples where the \( i \)th sample contains \( N_i \) observations and is characterized by a parameter, \( X_i \). The sample parameter, \( X_i \) approximates a parameter for the underlying population, \( \chi_i \). Suppose we wish to test the alternative hypotheses,

\[
\begin{align*}
H_0 & \quad \chi_i = \chi_j \quad \forall i, j, i \neq j, 1 \leq i, j \leq k \\
H_1 & \quad \chi_i \neq \chi_j \quad \exists i, j, i \neq j
\end{align*}
\] (2.1)

There are four possible outcomes of the test.

<table>
<thead>
<tr>
<th>( H_0 )</th>
<th>( H_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \chi_i = \chi_j )</td>
<td>( \chi_i \neq \chi_j )</td>
</tr>
</tbody>
</table>

Fail to Reject

- \( X_i = X_j \)
  - Correct
  - False Negative

Reject

- \( X_i \neq X_j \)
  - False Positive
  - Correct
We can use a statistical test, $F$, to examine the probability of a false positive, $p_a$. We evaluate $F(S_1, S_2, ... S_k) = a$, where $a$ is drawn from the $A$ distribution with parameters $c$. When $p_a$ is less than some predetermined critical value, $\alpha$, we reject $H_0$. The probability of a false negative is harder to measure by a direct test. This value is typically defined for each test. When the null hypothesis is true, $a$ is drawn from an $A$ distribution with parameters, $c$. If the alternative hypothesis is true, $a'$ should be drawn from an $A'$ distribution, characterized by parameters $c$ and a noncentrality parameter, $\lambda$ (70). The two-tailed probability of a false negative, $\beta$ is given by equation (2.2).

$$\beta = P[A' < a_{1-\alpha/2}] - P[A' < a_{\alpha/2}]$$

(2.2)

Statistical Power is defined as

$$PWR = 1 - \beta$$

(2.3)

which allows us to rewrite equation (2.2) as equation (2.4).

$$PWR = 1 - \beta$$

$$= 1 - (P[A' < a_{1-\alpha/2}] - P[A' < a_{\alpha/2}])$$

$$= 1 - P[A' < a_{1-\alpha/2}] + P[A' < a_{\alpha/2}]$$

$$= 1 - \Phi_{A'}(a_{1-\alpha/2}(c), c, \lambda) + \Phi_{A'}(a_{\alpha/2}(c), c, \lambda)$$

(2.4)

Where $\Phi_{A'}$ is the cumulative probability distribution function from the non-central $A$ distribution with noncentrality parameter, $\lambda$. The noncentrality parameter is a function of the effect size, which means a difference in the parameter, and the sample size (70).

2.2.1 Mathematical Definition of Empirical Power

Test-based power methods are a mainstay of traditional power calculations. However, the distribution and noncentrality parameter must be defined for these tests, which limits their applicability to nonparametric and permutation tests. We propose the use of Monte Carlo Simulations to ameliorate this problem.
Let us assume that we are testing the hypotheses described in equation (2.1). We know, or assume that for at least one pair of populations, \( P_i \) and \( P_j \), \( \chi_i \neq \chi_j \). When we apply a statistical test, \( G(S_1, S_2, ..., S_k) = p_{all} \) and let \( p_{all} \leq \alpha \), where \( \alpha \) is a threshold for rejecting \( H_0 \).

If \( N_i \) is sufficiently large, we can randomly subsample \( S_i \) to generate a subsample, \( s_{ij} \) of size \( n_{ij} \) with parameter, \( x_{ij} \), where \( n_{ij} < N_i \) and \( x_{ij} \approx X_i \), so by the transitive property, \( x_{ij} \approx \chi_i \). If we subsample \( m \) we can apply the the test, \( G \) and generate \( m \) test statistics, \( \{p_1, p_2, ..., p_m\} \).

Since we know that we should reject the null hypothesis, at the critical value, \( \alpha \), when \( p_j \geq \alpha \forall j | 1 \leq j < m \), we have committed a type II error. This lets us define the empirical probability of committing a type II error, \( \hat{\beta} \) according to equation 2.5.

\[
\hat{\beta} = \frac{|a_j > \alpha|}{m}
\]  

(2.5)

where \(|y|\) is the number of elements in set \( y \).

The empirical statistical power is be given by

\[
PWR = 1 - \hat{\beta}
\]

\[
= 1 - \frac{|a_j > \alpha|}{m}
\]

\[
= \frac{|a_j \leq \alpha|}{m}
\]  

(2.6)

As the number of subsamples performed, \( m \to \infty \) the central limit theorem, as applied to weak convergence theorem says that \( \hat{PWR} \to PWR \).

2.2.2 Pseudo Effect sizes

A second important feature of power calculations are the accompanying effect sizes. These can be estimated from the empirical power. We'll start by considering a test-based power, and the comparison of two means. Assume we have a two samples, each of size \( n \), drawn from defined populations with means \( \mu_1 \) and \( \mu_2 \) and the same variance, \( \sigma^2 \) where \( \mu_1 \neq \mu_2 \).

\[
H_0 \quad \mu_1 = \mu_2
\]

\[
H_1 \quad \mu_1 \neq \mu_2
\]  

(2.7)
We can test these hypotheses using a Z test. For a Z test, the test statistic, \( z \) is calculated as (71; 70)

\[
z = \sqrt{\frac{n}{2}} \left( \frac{\mu_1 - \mu_2}{\sigma} \right)
\]  

(2.8)

If the alternative hypothesis is true, the noncentrality parameter, \( \lambda_z \) is given by (70)

\[
\lambda_z = \left( \frac{\mu_1 - \mu_2}{\sigma} \right) \sqrt{\frac{n}{2}}
\]  

(2.9)

The power, as described by equation (2.4) can be calculated by substituting in equations (2.8), (2.9).

\[
PWR = 1 - \Phi_{Z'} (z_{1-\alpha/2}, \lambda_z) + \Phi_{Z'} (z_{\alpha/2}, \lambda_z)
\]

\[
= \Phi_Z (\lambda_z - z_{1-\alpha/2}) + \Phi_Z (z_{\alpha/2} - \lambda_z)
\]  

(2.10)

We can use equations (2.3), (2.9), and (2.10) to say that for a two-tailed test,

\[
1 - \beta = \Phi_Z \left( \sqrt{\lambda_z^2} - z_{1-\alpha/2} \right)
\]

\[
z_{1-\beta} = \sqrt{\lambda_z^2} - z_{1-\alpha/2}
\]  

(2.11)

Cohen (52) defined the standard effect size, \( d \) for a Z test as equation (2.12).

\[
d_z = \frac{\mu_1 - \mu_2}{\sigma}
\]  

(2.12)

We can substitute \( d_z \) into equation (2.11), which lets us approximate the effect size as a function of the power and number of observations.

\[
z_{1-\beta} + z_{1-\alpha/2} = \sqrt{d_z^2 \frac{n}{2}}
\]  

(2.13)

\[
d_z = \sqrt{\frac{2 (z_{1-\beta} + z_{1-\alpha/2})^2}{n}}
\]  

(2.14)
This allows an estimate of pseudo effect size, \( \hat{\delta} \), by substituting the empirical power, \( \hat{PWR} \) into (2.14).

\[
\hat{\delta} = \sqrt{\frac{2}{n} \left( z_{\hat{PWR}} + z_{1-\alpha/2} \right)^2}
\] (2.15)

As a caveat, the pseudo effect size is bounded such that the pseudo effect size can only be estimated when \( \hat{PWR} \) is between 0 and 1. \( z_x \to \infty \) as \( x \to 1 \), and \( z_x \to -\infty \) as \( x \to 0 \). Thus, effect size calculations may lose accuracy for very small (\( \hat{PWR} < 0.05 \)) or very large (\( \hat{PWR} > 0.99 \)) power estimates. In this paper, we have denoted effect sizes for very large power as \( \hat{\delta} \geq 4 \).

2.3 Implementation

2.3.1 Algorithm

In the most basic implementation, this algorithm subsamples an input matrix (samples), at a specified depth or depths (counts), and applies a test function (test). Power is calculated as the fraction of tests which return a p value less than the specified critical value (alpha). Figure 2.1 shows a basic implementation of the algorithm. Two special cases can be considered in this implementation. The first is a case of uneven sample size. The algorithm can be adjusted to accept a modifier (ratio) that allows consistent differences between the sample group. The second special case occurs when the observations in the samples matrix are matched, as might occur with a regression or correlation vector. Here, there are two options. The test function can be designed for the specific sample matrix, and a dummy index matrix could be subsampled. However, this means test must be defined uniquely every time, and relies on the user to know this procedure. A second, more robust solution is to include a variable, matched which specifies that matched samples should be drawn. We’ve chosen to implement the second solution. Both special cases are implemented in Appendix B.
Figure 2.1: The Monte Carlo Power estimation. (A) The inputs for the algorithm are `test`, statistical test which takes a partitioned set of samples and returns a p value; full lists of observations partitioned by group (in this case, solid or stripe), and a set of parameters including `depths`, or the number of observations which should be drawn as a subsample size; `num_runs`, the number of times power should be calculated at each depth; `iterations`, the number of tests which should be combined to form a power calculation; `alpha`, the critical value for power calculations. (B) Each iteration consists of subsampling the data to a specified depth, and applying `test` to the subsamples. (C) For each run (performed at the same depth), `iteration` iterations are performed, and the p values tabulated. The power estimate for the run (shown here as a barchart) is the number of p values less than `alpha`. (D) For each depth, a total of `num_runs` are performed, and used to tabulate the power. Power is then calculated at each depth to form the curve.
2.3.2 Metadata Pairing

In cases with complex data, it can be hard to deconvolve effect sizes from interactions. Nested studies and regression modeling can be useful to control for these variables, although regression is not widely used in microbiome data and is particularly challenging for beta diversity (71). We have developed a nesting schema within power studies (Appendix B). First, the observations are partitioned into groups based on control and test criteria. The smallest number of observations in each test group is determined for each combination of control groups. To draw a subsample, a number of observations to be drawn from each control group is selected. That number of observations is then drawn from each test group within the control, and the test is performed (Figure 2.2). This approach does not directly control for variability, but provides a limited measure of assurance that pairings are not exclusively driven by larger effects. It can also be used to test for interactions, where indirectly controlling for one effect may decrease the observed power of another.

2.3.3 Assumptions

Power and effect size calculations through Monte Carlo simulations make certain assumptions about the data. First, the samples are assumed to be random, and representative. Second, the method assumes that appropriate hypotheses and statistical tests have been selected. This includes passing along any assumptions associated with the statistical tests. When parametric tests are used, the assumptions of normality must hold. However, unlike traditional test-based methods, our empirical estimation can also be applied to permutative and nonparametric tests. Empirical power functions on the assumption that there is a significant difference between the underlying populations for the parameter being tested. It also assumes that the test can detect this difference. If the test is able to detect a significant difference between the full samples, this supports the assumption.
Figure 2.2: Model for subsampling based on metadata control categories. (A) In typical sampling, observations are partitioned based on a single property of the sample, in this case, whether they are solid or patterned. (B) Observations are then drawn at random from each category. (C) In metadata pairing, observations are partitioned based on whether or not they are solid or patterned (the test category), but also based on their shape. (D) Subsampling is performed by first determining a number of observations to draw from each control group, up to the minimum size of shapes in this category, and then drawing that number of observations from each treatment group in the control category. The subsample is then passed into the test, as shown in Figure 2.1B.
2.3.4 Parameter Optimization

One of the primary parameters to optimize for empirical power is the number of combinations required for estimation. This can be used to dictate a lower bound for sample size using power estimation. Data appropriate for a one sample t test was simulated with 1000 observations per group. Distributions were simulated by sweeping Cohen’s d-based effect sizes (0.1 to 1.4). For each distribution, power was calculated by drawing 5, 10, and 15 observations per sample. The number of subsamples per test were varied from 10 to 5000, and the number of tests combined was varied between 1 and 25. Power was calculated from the \( \hat{\delta} \) at each point. The range of effect sizes observed and the range of power values for 25 observations in the sample with a critical value of 0.05 were used as guides to determine the minimum number of iterations required (Figure 2.3).

The simulations suggest that at lower effect sizes, 1000 iterations per run, with 25 runs combined, are required so the power range is 0.01, even at lower effect sizes. The runs may be combined over multiple count depths, so if 10 runs are performed per depth at 3 to 5 depths, the power will collapse within a reasonable range, even at low effect sizes. If we are able to draw 3 observations, we need about 9 runs per depth, or at least 9000 possible combinations per depth. If we can draw 5 depths, we need 5 runs per depth, and 5000 possible combinations per depth.

Let us assume that under most circumstances, the minimum subsample size is five observations. This is premised on the rule of thumb that says rank-sum tests should not be attempted with less than 5 observations per group. The number of possible combinations of size \( r \) which can be drawn from a sample of size, \( n \) is \( nC_r \), defined by equation (2.16).

\[
nC_r = \frac{n!}{r!(n-r)!}
\]  

(2.16)

If we draw \( r_i \) observations from ith independent sample of size \( n_i \), we can multiply the combinatorial space associated with each of the the \( k \) samples.

\[
C_T = \prod_{i=1}^{k} n_iC_{r_i}
\]

(2.17)

Metadata adjustment typically results in a smaller overall sample, since samples which do not
Figure 2.3: Power extrapolated from pooled $\hat{\delta}$ collapses to the mean power as a function of the number of tests and number of power calculations combined. Power was simulated by varying the number of tests used to calculate power between 10 and 5000 and the number of power calculations combined between 1 and 10. (A) An example of the power spread for an effect size of approximately 0.25. The red line represents the mean power for the simulation. The log$_{10}$(range) of the (B) mean extrapolated effect size and (C) predicted power are shown for 1 (○), 3 (▲), 5 (★), and 10 (□) observations combined as a function of effect sizes between 0.1 and 1.25.
fully match the rest of the data are excluded (Figure 2.2). However, paired subsamples still allow for a large combinatorial space. Functionally, we can treat the partitioning as a multi-factorial design, where we have $I$ control groups and $J$ treatments. The samples in each of the treatments are independent. The $i$th control group for the $j$th treatment has $n_{ij}$ observations. We can draw at most $r_i$ observations from the $i$th control group, where

$$r_i = \min(n_{ij}) \quad (2.18)$$

In total, we can draw $R$ observations.

$$R = \sum_{i=1}^{I} r_i \quad (2.19)$$

Let us assume we plan to draw $Q$ observations in total ($1 \leq Q \leq R$) We will draw $q_i$ observations ($p_i \leq q_i \leq r_i$) from each group.

$$p_i = \min(\max(0, Q - \sum r_i), r_i) \quad (2.20)$$

$$Q = \sum_{i=1}^{I} q_i \quad (2.21)$$

For some solution $Q_x$, ($\{q_{x1}, q_{x2}, \ldots q_{xl}\}$), there are $n_{ij}C_{q_{xi}}$ ways to draw $q_{xi}$ samples from the $i$th control group and the $j$th treatment. Each of the treatment groups are independent, so can multiple their combinations. The number of possible combinations for the $x$th solution and the $i$th treatment group, $C_{xi}$ is

$$C_{xi} = \prod_{j=1}^{J} n_{ij}C_{q_{xi}} \quad (2.22)$$

Once the group sizes are set, the ways the sample are drawn for each observations are independent. So, the possible number of combinations are
\[ C_x = \prod_{i=1}^{l} C_{x_i} \]
\[ = \prod_{i=1}^{l} \prod_{j=1}^{J} n_{ij} C_{q_{xi}} \]  

(2.23)

Now, let's turn our attention to how we optimize the \( q_{xi} \) space. There are a finite number of values of \( q_i \) which will satisfy the limits imposed by equations 2.18, 2.20 and 2.21. These create boundary conditions where

\[ p_1 \leq q_1 \leq r_1 \]
\[ p_2 \leq q_2 \leq \min (Q - r_1, r_2) \]
\[ p_i \leq q_i \leq \min \left( Q - \sum_{x=1}^{i} r_x, r_i \right) \]  

(2.24)

We'll call the upper limit for each iteration \( m_i \). To get the total number of combinations, \( C_{all} \), we sum over the \( i - 1 \) boundary conditions where the number of samples are not fixed. For example, for a model where the samples are divided between two control groups, the combinatorial space would be

\[ C_{all} = \sum_{q_1=p_1}^{r_1} \prod_{j=1}^{J} \left( n_{1j} C_{q_1} \right) \left( n_{2j} C_{Q-q_1} \right) \]  

(2.25)

For three groups, the total combinatorial space would be

\[ C_{all} = \sum_{q_1=p_1}^{r_1} \sum_{q_2=p_2}^{r_2} \prod_{j=1}^{J} \left( n_{1j} C_{q_1} \right) \left( n_{2j} C_{q_2} \right) \left( n_{3j} C_{Q-q_1-q_2} \right) \]  

(2.26)

Figure 2.4 explores the combinatorial space for samples of varying sizes, and drawing varying number of observations using traditional sampling. For a single sample, we can achieve the necessary number of combinations if we sample depths of 5, 10, and 15 observations from a total sample size of 20 observations. A group size of 16 will satisfy the same conditions for two samples of equal sizes and equal depths. However, it can be wise to err on the side of more samples in each group, since more samples increase the probability of finding a significant difference. With this in mind, it
Figure 2.4: Combinatorial depth as a function of sample and subsample size. Observations were drawn from (A) one sample (B) two samples of equal size and (C) two samples with unequal sample sizes. Images are colored as the log of the number of combinations. A single sample has the smallest number of combinations, although a sample size of 20 observations allows appropriate combinatorial space.

is generally advisable to design de novo experiments with at least twenty observations per sample for each sample being tested.

2.4 Validation

2.4.1 Pseudo effect sizes

The pseudo effect size, $\hat{\delta}$ was validated in two ways. Test based power was simulated for data appropriate to a one sample t test, two sample t test, one way analysis of variance, and Pearson’s correlation coefficient. The $\hat{\delta}$ effect size was calculated using the test based power using 100 simulations. Power was then estimated based on the effect size. If we averaged over all power values, the correlation coefficient was $R^2 = 0.994$, and the ratio was $2.6 \pm 0.6$ (mean $\pm$ 95% CI) (Figure 2.5A). When we removed observations where the traditional power was less than 0.05 (the critical value used to calculate power), the correlation coefficient was 0.995 and the ratio was 0.997 $\pm$ 0.002. The effect-size based power overestimates power for all five tests at lower power values and lower effect sizes. With that caveat, $\hat{\delta}$ is an accurate way to extend power.

The pseudo effect size was also compared to effect size estimators published by Cohen (52). The effect sizes were evaluated separately for one sample t tests, independent sample t tests,
Table 2.1: Validation of Cohen’s d, standardized pseudo δ and adjusted pseudo δ for effect sizes between 0.1 and 1.5

<table>
<thead>
<tr>
<th>Test</th>
<th>Adjustment</th>
<th>$R^2$</th>
<th>Standardized Ratio</th>
<th>Adjusted Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent Sample T Test</td>
<td>1</td>
<td>0.999</td>
<td>1.112 [1.105, 1.119]</td>
<td>1.112 [1.105, 1.119]</td>
</tr>
<tr>
<td>One Sample T Test</td>
<td>$\sqrt{2}$</td>
<td>0.997</td>
<td>0.763 [0.760, 0.766]</td>
<td>1.079 [1.075, 1.083]</td>
</tr>
<tr>
<td>One way ANOVA</td>
<td>2</td>
<td>0.999</td>
<td>0.571 [0.569, 0.573]</td>
<td>1.141 [1.136, 1.146]</td>
</tr>
</tbody>
</table>

and one way analysis of variance (ANOVA). There is a constant modifier, $c_{test}$ which maps $\delta$ to Cohen’s effect size (Table 2.1; Figure 2.5B). The constant values were based on the test, and the relationship between the noncentrality parameter for the parametric test, $\lambda$, and the effect size. The noncentrality parameter for a one sample t test is given by equation (2.27).

$$
\lambda_{t1} = \frac{\bar{x} - \mu}{\sqrt{\frac{ns}{2}}} \quad (2.27)
$$

The base effect size used to calculate the $\delta$ is given in equation (2.9). The relationship between these two effect sizes is $\lambda_{t1}\sqrt{2} = \lambda_Z$. Similarly, the noncentrality parameter for an ANOVA with $k$ treatments is

$$
\lambda_f = \sum_{i=1}^{k} \lambda_i^2
$$

$$
= \sum_{i=1}^{k} \left( \frac{\bar{x}_i - \bar{x}_{..}}{\sqrt{\frac{n_i}{s}}} \right)^2 \quad (2.28)
$$

The relationship between this noncentrality parameter and $\lambda_z$ is 2. So, we will multiply the $\delta$ by 2 for an ANOVA. Once corrected, the combined tests had a correlation coefficient $R^2 = 0.990$, with a ratio of 1.074 ± 0.007. Most of the errors came from effect sizes where the Cohen-based effect sizes ($d$ or $f^2$) were less than 0.1 or greater than 1.5. Given the power estimates above were not accurate for lower powers, and that the effect size cannot accurately be estimated using this method once power is 1, it is unsurprising that extremely low or extremely high effect sizes cannot be estimated using this method. Once the correction was implemented, the pooled $R^2$ was 0.998, and the ratio was 1.110 ± 0.003 (Table 2.1).
Figure 2.5: A pseudo effect size can accurately return power and Cohen's coefficients. Power was calculated using test-based methods on simulated normal data. The pseudo effect size was calculated from the test-based power using Equation (2.15); extrapolated power was calculated according to (2.11). (A) The extrapolated power is highly correlated with the test-based power for power values greater than 0.05, the critical value. (B) There is a linear relationship between effect sizes and Cohen's $f^2$ for an ANOVA (○), Cohen’s $d$ for a one sample $T$ test (□) and Cohen’s $d$ for a two sample independent $T$ test (◆). The effect sizes shown here have not been corrected.
2.4.2 Application to normal data and test based methods

We simulated data appropriate for testing with a one sample T test, two sample T test, one way ANOVA, and Pearson’s correlation coefficient (70). Power was tested using test-based methods, the empirical method, and from an extrapolated effect size (Table 2.2). The test-based power was highly correlated with the empirically calculated result \((R^2_{pooled} = 0.991)\), and the ratio was 1.21 ± 0.02. Power was also calculated using the pseudo-effect size, and compared to the test-based method \((R^2_{pooled} = 0.990)\), with a ratio of 1.010 ± 0.004. Both the empirical and extrapolated methods slightly overestimated the power, compared to the test-based methods (Figure 2.6A, C).

The pseudo effect size was also examined for T tests and ANOVA (Table 2.2; Figure 2.6B). Effect sizes were modified by the constants from Table 2.1. Pooling the effect sizes, the correlation between Cohen’s effects and the empirical \(\hat{\delta}\) was 0.979, and the ratio 1.14 ± 0.01. The \(\hat{\delta}_{test}\) from the test-based power was also compared to the empirical effect \(\hat{\delta}_{empirical}\). The correlation was 0.972, and the ratio 1.09 ± 0.01. Plotting the curve suggested the effect size estimation is inaccurate for traditional power values less than 0.2 or greater than 1.5. When these samples were removed, the correlation coefficient between Cohen’s effect size and \(\hat{\delta}_{empirical}\) was 0.978, with a ratio of 1.07 ± 0.01. In comparison, the correlation between \(\hat{\delta}_{test}\) and \(\hat{\delta}_{empirical}\) was 0.972, and the ratio was 1.047 ± 0.007 (Figure 2.6D).

The comparison between the empirical and test-based methods demonstrate that power estimated by the empirical methods is a good approximation of power calculated using test-based techniques. Power can be accurately extrapolated using the pseudo effect size, calculated from empirical results. There is a linear relationship between the pseudo effect sizes and the test-based effect sizes proposed by Cohen. This suggests empirical power is both accurate and useful for power and effect size calculations.
Monte Carlo based power is able to accurately return test-based power for normal data. There is a linear relationship between test-based methods, (A) Monte Carlo estimation, and (B) extrapolated form a calculated \( \hat{\delta} \). (C) There is a linear relationship between Cohen’s \( f^2 \) for an ANOVA (\( \circ \)), Cohen’s \( d \) for a one sample T test (\( \square \)) and Cohen’s \( d \) for a two sample independent \( t \) test (\( \star \)) and the \( \hat{\delta} \) for each test. (D) Test-based power and empirical power are related through extrapolation of \( \hat{\delta} \).

### Table 2.2: Pseudo effect size (\( \hat{\delta}_{c_{mod}} \)) from empirical power for cohen effects between 0.2 and 1.5

<table>
<thead>
<tr>
<th>Test</th>
<th>Cohen’s effect</th>
<th>Test-based ( \delta )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( R^2 )</td>
<td>( R^2 )</td>
</tr>
<tr>
<td>Independent Sample T Test</td>
<td>0.993 1.13 [1.11, 1.15]</td>
<td>0.993 1.06 [1.05, 1.07]</td>
</tr>
<tr>
<td>One Sample T Test</td>
<td>0.979 1.136 [1.129, 1.143]</td>
<td>0.969 1.106 [1.098, 1.114]</td>
</tr>
<tr>
<td>One way ANOVA</td>
<td>0.985 0.999 [0.993, 1.005]</td>
<td>0.985 0.901 [0.896, 0.906]</td>
</tr>
</tbody>
</table>
2.4.3 Applications to non-normal data

We tested the performance of Monte Carlo Power on non-normal data. We simulated two bimodal distributions using a binomial random variable with the same shape and scale parameters. The distributions were offset from each other by some constant, $\Delta$, between 1.5 and 3.5. The means of distributions were compared using a parametric independent sample T test and a non-parametric Mann Whitney U test. Power was calculated using a test-based method (70) and with our Monte Carlo power estimate. Cohen’s d and the $\hat{\delta}$ were also calculated. Test-based methods underestimated the power and effects for comparisons of binomial distributions (Figure 2.7). Therefore, we have demonstrated that Monte Carlo power estimation is more appropriate for non-normal data than a test-based method.

2.4.4 Power extrapolation with underpowered datasets

Studies, especially pilot studies, may not be powered, but can still demonstrate significance. Therefore, it is important the $\hat{\delta}$ effect size can be calculated on underpowered data. To explore this question, we simulated normal and nonnormal distributions with 125 observations in each distribution. We then subsampled these distributions to give us samples with 25 observations. We calculated Cohen’s d and $\hat{\delta}$ for the full sample and subsample (Figure 2.8). We compared effect sizes from samples where all effect sizes were greater than 0.1. The difference between the 125 observation sample and the 25 observation subsample appears to be a subsampling bias, as the differences are seen both in Cohen’s d and $\hat{\delta}$. This means that $\hat{\delta}$ can be used on samples which are not fully powered, and behaves as well as test-based effect sizes.

2.5 Application to Microbiome Studies

Several studies have been used to establish the role of body site, age, nationality, IBD, and BMI status on the microbiome. Given the study size requirements, we elected to examine the biological effects of body site in adults (29) based on the Human Microbiome Project (HMP);
Figure 2.7: Monte Carlo based power handles nonparametric data better than test-based power. Two binomial distributions were simulated by combining two gamma random variables, offset from each other. (A) Power was calculated using test-based methods for an independent sample T test (70), using Monte Carlo simulations with a Mann Whitney U test (“Empirical Power”) and from the $\tilde{\delta}$ calculated from the Monte Carlo power (“Extrapolated Power”). (B) At the same sampling depths (10 observations per group), the test-based method gave a lower effect size (black) than the Monte Carlo method (blue).
Figure 2.8: Subsamples of the data are able to recapitulate effect sizes. We calculated power for the comparison of 125 samples drawn from (A) two normal distributions or (B) two binomial distributions. (C, E) A subsample of 25 were drawn from the full data set. Power was calculated using test-based and Monte Carlo methods for for full dataset (A, C) and the subset (B E). We compared the ratio between the $\hat{\delta}$ and $d$ for the all observations and the subset. The parametric data (C) had a ratio of 1.17 [1.08, 1.25] (95% CI), and (F) the nonparametric data had a ratio of 0.89 [0.88, 0.99] (95% CI). (G, H) We also compared the calculated Cohen’s $d$ and $\hat{\delta}$ for all observations and the subset of observations. The difference between the $d_{subset}$ and $d_{all}$ was the same as the relationship between the $\hat{\delta}_{subset}$ and the $\hat{\delta}_{all}$. 
Table 2.3: Effect size associated with technical and biological variables

<table>
<thead>
<tr>
<th>Study</th>
<th>Reference</th>
<th>Variable</th>
<th>Standard $\hat{\delta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Microbiome Project</td>
<td>(29)</td>
<td>Body Site</td>
<td>$2.52 \ [2.49, 2.55]$</td>
</tr>
<tr>
<td>Yatsunenko et al</td>
<td>(30)</td>
<td>Age</td>
<td>$2.30 \ [2.25, 2.34]$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nationality</td>
<td>$2.01 \ [1.96, 2.05]$</td>
</tr>
<tr>
<td>Obesity Meta Analysis</td>
<td>(27)</td>
<td>Study $\geq 4^1$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Body Mass</td>
<td>$0.84 \ [0.82, 0.86]$</td>
</tr>
<tr>
<td>IBD Meta Analysis</td>
<td>(27)</td>
<td>Study</td>
<td>$3.28 \ [3.21, 3.34]$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IBD Diagnosis</td>
<td>$1.42 \ [1.39, 1.45]$</td>
</tr>
</tbody>
</table>

$^1$Power was 1 for all depths and $p$ values greater than the permutative $p$

age and Western or non-Western diets from Global Gut (GG), represented by nationality (30); Inflammatory Bowel Disease (IBD) (27; 72; 73; 74; 75); and Obesity (27; 29; 30; 74; 76; 77). The HMP and GG were trimmed to 100 nucleotides and picked closed reference against August 2013 Greengenes release using SortMeRNA in QIIME 1.9 (78; 79; 80). The Obesity and IBD meta-analysis were picked closed reference against the August 2013 Greengenes release using UClust and QIIME 1.8 (78; 80; 81). All samples were rarefied to 1000 sequences per sample, and studies were compared using a permanova applied to unweighted UniFrac distance (65; 68). Data were examined visually using principal coordinates analysis in Emperor (82). Initially, we calculated power on all groups in the category without filtering any variables. Using 999 permutations, all variables were significant at $p = 0.001$. The Permanova test does not follow formal requirements for test-based power because data are not normal and no assumptions are made about the underlying distribution of either the data or the test statistic (68). Despite equal permutative $p$ values, the observed effect sizes varied widely, and reflected previously observed biological conclusions (Table 2.3, Figure 2.9) (28).

The technical variation due to study-to-study differences had the largest effect on community structure. The strength of study effects have been reported in previous meta-analyses, although they are generally considered on the same order as biological effects in adults, such as body sites (28). Meta analyses in the microbiome field therefore typically process the data through the same data analysis pipeline, but focus comparisons within studies. We further investigated study effect
Figure 2.9: Effects associated with the human microbiome. (A) We calculated power curves for the IBD study (yellow), IBD status (gray), obesity study (pink), obesity BMI (green), HMP body site (teal), age (purple), and nationality (orange). The inset shows power across the first five observations, due to the large effect sizes. We also examined principle coordinates projections of unweighted UniFrac distance for (B) IBD study, (C) IBD disease status, (D) Obesity study, (E) BMI, (F) Obesity study and BMI, (G) body site, (H) age, and (I) nationality.
in the lean and obese cohorts from Walters et al (27). Three of the studies analyzed amplified the same hyper-variable region, V2, and used the same sequencing method. We compared the study effect in adults twenty years and older for all the projects in the study and the projects which used the V2 region (76; 77; 74) (Figure 2.10A, B). The study effect was greater than 4, as we could not find a reasonable sub sampling depth which optimized both permutative space associated with power calculations and permanova, but returned a result less than 1 in any subset of the data involving multiple variable regions. The V2 study comparison had an effect size of 2.5 [2.2, 2.8] (Figure 2.10).

We used fecal samples from the Human Microbiome Project to estimate the effects associated with different primers and processing centers. Two hyper variable regions (V1-3 and V3-5) were used, and samples were processed at four different sites. We examined the effect of the primer across all sites. We also compared the effect of processing between two sites using samples amplified using the V3-5 hyper variable region (Figure 2.10C, D). The effect of varying the region was 1.5 [1.3, 1.6], while the center effect was 0.32 [0.28, 0.36]. The PCoA space reflects the effect size scale: The study effect drove clustering in the PCoA comparing the HMP to other studies of lean and obese adults (27) (Figure 2.10A), although the primer effect was stronger than anything else observed in the HMP alone (Figure 2.10D). In comparison, the largest biological effect associated with obesity within each study was 0.25 ([0.2, 0.3]).

2.6 Prospectus

These results represent the first comparison of quantified study effects to quantified biological effects. We replicated results demonstrating the study effect is stronger than biological effects (27; 28). However, unlike previous comparisons which relied on separations in PCoA space to assign effects, we were able to calculate effects on variables, individually. The results strongly support calls for technique standardization, including the adoption of uniform primer regions across a biome, and regular lab-to-lab validation, to control for processing center effects.

We have developed and validated a new method for estimating statistical power using Monte
Figure 2.10: Study effects have a larger impact on microbial communities than biological effects. (A) Statistical power for the study effect, and primer effect are much larger than the effect of processing at a different location or the largest within-study effect due to obesity. The difference in power is reflected in the principal coordinates plots, where separation is based on study in the (B) Walters’ obesity studies for all studies and (C) the studies with V2 primers alone. Within the studies, the BMI of the participant does not drive clear separation, (D) The primer choice drives separation within fecal samples from the Human Microbiome Project. (E) Although we cannot see clear separation by sample site, there is a significant difference (p < 0.001) between samples processed at the Broad and JCVI.
Carlo simulation. Our new method performs as well as test-based approaches on parametric data, and is able to recapitulate Cohen’s effect sizes using a constant modifier. The Monte Carlo estimation out performs test-based power when applied to nonparametric data. Furthermore, the Monte Carlo method can be applied to data where test based methods are not possible, because of the lack of knowledge \textit{a priori} about the underlying distribution, including Permanova. The extrapolated \( \hat{\delta} \) effect size is as disposed to sampling bias as test-based statistics, and can be used to extend statistical power. Unlike previous methods to establish power, ours is test-agnostic. We applied this method to measure study effects from a previously published meta-analysis (27). We demonstrated that study effects are often stronger than biological effects, but that there are detectable differences due to biology, as well. As we have seen, calculating statistical power for microbiome studies is challenging but addressable in a Monte Carlo context. In the next chapter, I apply this technique to the American Gut cohort, the largest crowdfunded citizen-science initiative to date.
American Gut: an Open Platform for Citizen-Science Microbiome Research

In the previous two chapters, I described the potential of participatory microbiome research, and introduced a new method to calculate effect size and statistical power for microbial data. In this chapter, I apply these techniques to the American Gut cohort. American Gut is the largest crowdfunded, crowdsourced, microbiome project that we are aware of, covering thousands of people (compared to the hundreds in previous large-scale projects including the NIH-funded Human Microbiome Project). The large number of people involved provide a unique opportunity to identify which factors produce major changes in the microbiome, and which are more subtle, potentially guiding a wide range of future studies. This work is based on paper in preparation for submission to Science, where my specific contributions included leading the primary analysis of the American Gut data, assisting in the development of primary processing and results generation pipeline, web development, work on participant interaction, and co-leading the Consortium effort to write this manuscript.

The American Gut Project was launched in November 2012 to characterize the diversity of the human microbiome in people regardless of lifestyle. The new reference database of human microbiome samples allowed us to characterize the diversity of the North American gut microbiome, describe correlations with participant health, lifestyle, and diet, and establish the American Gut resource and infrastructure as a platform for discovery (for example, through targeted sub-studies that are distinguished against the full dataset).

As of March 31st, 2015, the American Gut Project includes microbial sequence data from
Table 3.1: Summary of samples by sample type in the American Gut (AG) as of May 27th, 2015 compared with the Human Microbiome Project (HMP)

<table>
<thead>
<tr>
<th>Sample Type</th>
<th>AGP Samples</th>
<th>AGP Participants</th>
<th>HMP Samples</th>
<th>HMP Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fecal</td>
<td>3679</td>
<td>3338</td>
<td>365</td>
<td>230</td>
</tr>
<tr>
<td>Skin</td>
<td>336</td>
<td>178</td>
<td>1367</td>
<td>238</td>
</tr>
<tr>
<td>Oral</td>
<td>358</td>
<td>327</td>
<td>3316</td>
<td>234</td>
</tr>
<tr>
<td>Vaginal</td>
<td>8</td>
<td>7</td>
<td>482</td>
<td>109</td>
</tr>
<tr>
<td>Nasal</td>
<td>6</td>
<td>6</td>
<td>339</td>
<td>221</td>
</tr>
<tr>
<td>Hair</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Blank</td>
<td>435</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

5,020 samples from 4,199 human participants, whose microbes are represented by over 130 million 16S rRNA gene fragments from variable region 4. The primary specimen type is fecal, totaling 4,279 samples from 3,889 participants. Skin and oral body sites are also represented, totaling 343 and 368 samples, respectively (additional detail in Table 3.1). The age range represented spans infants less than a year old to the elderly up to 93 years of age, body mass indices from 8.5 to 78.1, and a multitude of other health and lifestyle variables. Only individuals with felony convictions and those younger than six weeks old are excluded from the study due to Institutional Review Board restrictions on research in these protected populations. The project is crowd-funded, initially through contributions to IndieGogo (83) and currently through contributions through the Fundrazr portal (84). A demographic breakdown of the participants can be found in Table reftab:agS3. Participants span urban and rural boundaries, race, and ethnicity in greater numbers than found in the Human Microbiome Project (HMP), a cohort of 242 individuals-mainly in their 20s and 30s-from St. Louis, MO and Houston, TX (29). So far, the majority of the samples collected were from the Caucasians living in the continental United States.
Table 3.2: Demographics of US participants in the American Gut

<table>
<thead>
<tr>
<th></th>
<th>Counts</th>
<th>Percentage*</th>
<th>Published</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1828</td>
<td>55.1</td>
<td>50.8</td>
</tr>
<tr>
<td>Male</td>
<td>1487</td>
<td>44.7</td>
<td>49.2</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>0.03</td>
<td>—†</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>24</td>
<td>0.67</td>
<td>12.6</td>
</tr>
<tr>
<td>Asian or Pacific Islander</td>
<td>3</td>
<td>80.5</td>
<td>0.</td>
</tr>
<tr>
<td>Caucasian</td>
<td>3317</td>
<td>92.0</td>
<td>63.7</td>
</tr>
<tr>
<td>Hispanic‡</td>
<td>55</td>
<td>1.53</td>
<td>16.3</td>
</tr>
<tr>
<td>Other</td>
<td>72</td>
<td>2.00</td>
<td>6.2</td>
</tr>
<tr>
<td><strong>Body Mass Index§</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underweight</td>
<td>118</td>
<td>3.81</td>
<td>1.8</td>
</tr>
<tr>
<td>Normal</td>
<td>1917</td>
<td>62.0</td>
<td>31.2</td>
</tr>
<tr>
<td>Overweight</td>
<td>706</td>
<td>22.8</td>
<td>34.0</td>
</tr>
<tr>
<td>Obese</td>
<td>351</td>
<td>11.4</td>
<td>33.0</td>
</tr>
<tr>
<td><strong>Smoker</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I do not smoke</td>
<td>3417</td>
<td>95.6</td>
<td>82.6</td>
</tr>
<tr>
<td>I smoke</td>
<td>158</td>
<td>4.42</td>
<td>17.4</td>
</tr>
<tr>
<td><strong>IBD</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I do not have IBD</td>
<td>3300</td>
<td>94.9</td>
<td>99.6</td>
</tr>
<tr>
<td>I have IBD</td>
<td>176</td>
<td>5.06</td>
<td>0.4</td>
</tr>
<tr>
<td><strong>Diabetes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I do not have diabetes</td>
<td>3429</td>
<td>97.2</td>
<td>90.7</td>
</tr>
<tr>
<td>I have diabetes</td>
<td>97</td>
<td>2.75</td>
<td>9.3∥</td>
</tr>
</tbody>
</table>

Calculated as the size of the reported group / total reported
†Not reported.
‡The US census reports Hispanic as an ethnicity, not a race
§BMI categories are for participants older than 20
∥Diagnosed and undiagnosed
3.1 Materials and Methods

3.1.1 Participant Recruitment and Sample Processing

Participants signed up for the project through and Indiegogo (83) and Fundrazr (84). A monetary contribution to the project was made to help offset the sequencing costs. All participants were consented under an approved IRB protocol, either from the University of Colorado, Boulder (December 2012 - March 2015) or the University of California, San Diego (March 2015-present). The IRB document specifically allows for public deposition of all data that is not personally identifying and for return of results to participants.

Self-reported metadata was collected through a web portal (85). Samples were collected using BBL Culture Swabs (Becton, Dickinson and Company; Sparks, Maryland) and returned by mail. Samples were processed using the Earth Microbiome Project (EMP) protocols. Briefly, the V4 region of the 16S rRNA gene was amplified with barcoded primers and sequenced as previously described (86).

3.1.2 Primary Processing

As a subproject of the Earth Microbiome Project (EMP) (41), the American Gut Project is committed to open-access policies and minimizing barriers to reproducing analyses and access to data. Specifically, all sequence and de-identified participant data generated by the American Gut are deposited into EBI (87), and how the data are processed is written up in a BSD-licensed executable IPython Notebook that is part of the American Gut Github repository (88). The primary processing of the data, which produces the participant results and the raw OTU tables, is centered around QIIME (78). In brief, the Notebook performs the following steps: consumes data from EBI; removes bloom sequences; picks OTUs using SortMeRNA (79; 89) against the August 2013 release of Greengenes (80); produces taxonomy summaries; visualizes principal coordinates from UniFrac distances using EMPeror (82); and generates deliverable results for participants (Figure 3.1). All aspects of the data processing use open-source programs that are either GPL- or BSD-compatible.
Figure 3.1: An example of a participant result. The bar chart shows a phylum level taxonomic summary comparing the participant result to the overall American Gut cohort, to people of similar diet, BMI, gender, age, and the microbiome of food writer Michael Pollan. The tables describe the most abundant genera and families in the individual’s sample, as well as taxa enriched in the sample compared to the rest of the American Gut participants. Rare taxa were defined as those found in less than 10% of the American Gut population. The PCoAs across the bottom show the participant’s sample in reference to body site, fecal samples within the global gut, and the American Gut population alone. Results are regenerated as new data are added.
3.1.3 Bloom Filtering

Mailing contributed to overgrowth of certain bacterial strains in a subset of the American Gut samples. A subset (1210 samples) of the American Gut data were compared to 438 samples that were frozen immediately after collection from Personal Genome Project (PGP) participants. The comparison indicated that American Gut fecal samples contained higher abundances of Gammaproteobacteria (Figure 3.2). This observation, together with data from the Mayo fecal stability study and ongoing in-house methods validation suggest the differences in abundance are due, in large part, to a few taxa. These were identified as taxa which were present in the American Gut samples, but not freshly frozen samples from the PGP, Mayo fecal stability study, or Microbiome Quality Control Study (MBQC); taxa which were present in samples from the Mayo fecal stability study after four days at room temperature which were not found in the freshly frozen samples, and Gammaproteobacteria bacteria present in the American Gut samples at more than 3%. The list of sequences included nine Gammaproteobacteria and one Bacillus (Table 3.3). While these taxa could also be present in fresh samples, their sequences are filtered out as their high abundances are highly dependent on shipment and therefore can represent a large source of noise during downstream data analysis. The filtering procedure (manuscript in preparation) reduced the Gammaproteobacteria 6.5-fold in the American Gut samples, without significantly altering the abundance in the PGP samples. The taxonomic composition of both studies was highly correlated before and after filtering, as measured by a spearman rank correlation coefficient (r=0.9, permutative p = 0.001; Figure 3.2).

3.1.4 Statistical Analysis

Population level comparisons were calculated for all American Gut participants living in the United States. BMI categorization was only considered for adults over the age of twenty, since the description of BMI in children is based on their age and sex. The percentage of the American Gut participants was calculated as the fraction of individuals who reported results for that category. US
Figure 3.2: Bloom Filtering primarily alters the relative abundance of Gammaproteobacteria. There was, on average, a 6.5 fold reduction in Gammaproteobacteria in the mailed American Gut samples, compared to the immediately frozen PGP samples. The correlation between taxon abundance in filtered and unfiltered samples were 0.92, 0.91, and 0.90 respectively for American Gut sequencing runs 1, 2, and 7 and 0.92 for the PGP samples (all permutative p < 0.001). However, filtering reduced the Gammaproteobacteria abundance in the American Gut samples by 6.5 fold.
<table>
<thead>
<tr>
<th>Sequence ID</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>3007971-Bloom</td>
<td>TACGTAGGTGGCAAGGTTGTCGCAGAATTATTGGGCGTAAAGCGCGCGCAGGGCTCTTTTAAGTCTGATGTGAAGCCACGGCTCAACCGTGAGGGTCATTGGAAACTGGAAGACTTGAGTGCAGAAGAGAGTGGAATCCAGG</td>
</tr>
<tr>
<td>4459942-Bloom</td>
<td>TACGTAGGTGGCAAGGTTGTCGCAGAATTATTGGGCGTAAAGCGCGCGCAGGGCTCTTTTAAGTCTGATGTGAAGCCACGGCTCAACCGTGAGGGTCATTGGAAACTGGAAGACTTGAGTGCAGAAGAGAGTGGAATCCAGG</td>
</tr>
<tr>
<td>6359652-Bloom</td>
<td>TACGGAGGGTGCAAGGTTAATCGGAATTACTGGGCGTAAAGCGCRGCACTGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCCGAAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGGAATCCAGG</td>
</tr>
<tr>
<td>2376152-Bloom</td>
<td>TACGGAGGGTGCAAGGTTAATCGGAATTACTGGGCGTAAAGCGCRGCACTGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCCGAAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGGAATCCAGG</td>
</tr>
<tr>
<td>5235310-Bloom</td>
<td>TACGGAGGGTGCAAGGTTAATCGGAATTACTGGGCGTAAAGCGCRGCACTGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCCGAAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGGAATCCAGG</td>
</tr>
<tr>
<td>9016203-Bloom</td>
<td>TACGGAGGGTGCAAGGTTAATCGGAATTACTGGGCGTAAAGCGCRGCACTGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCCGAAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGGAATCCAGG</td>
</tr>
<tr>
<td>8491357-Bloom</td>
<td>TACGGAGGGTGCAAGGTTAATCGGAATTACTGGGCGTAAAGCGCRGCACTGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCCGAAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGGAATCCAGG</td>
</tr>
<tr>
<td>1963084-Bloom</td>
<td>TACGGAGGGTGCAAGGTTAATCGGAATTACTGGGCGTAAAGCGCRGCACTGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCCGAAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGGAATCCAGG</td>
</tr>
<tr>
<td>9894753-Bloom</td>
<td>TACGGAGGGTGCAAGGTTAATCGGAATTACTGGGCGTAAAGCGCRGCACTGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCCGAAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGGAATCCAGG</td>
</tr>
</tbody>
</table>
population data is based on the 2010 census, and Center for Disease Control reports on Diabetes, IBD and smoking (90; 91; 92; 93; 94; 95).

OTU tables generated in QIIME by closed-reference OTU with SortMeRNA against Green-genes (August 2013) by the primary processing notebooks (78; 79; 80). The OTU table was rarefied to 10,000 sequences per sample, and split by body site. Shannon, Chao1, Observed OTU and PD whole tree diversity were calculated as the mean of ten rarefactions in QIIME (96; 97; 13; 98). A single sample from each participant at each body site was selected at random; American Gut participants were able to submit multiple samples from multiple body sites. For all reported analyses, we randomly selected a single sample from each participant at each body site. We did this to control for autocorrelation, which has been previously observed across body sites and across time within individuals (99; 100) including in our own investigations of participants who submitted multiple samples (data not shown). This random participant sample set is consistent across all analyses with the exception of the study-wide results shown in Figure 3.3. Healthy adults were defined as individuals 20-69 with a BMI between 18.5 and 30 who did not report antibiotic use in the past year, and had no reported history of IBD or diabetes. Alpha diversity for single metadata categories were compared with a Kruskal-Wallis test. Unweighted UniFrac distance between samples was tested with permanova and permutative t-tests in QIIME (68; 78). The alpha diversity regression was performed using the Statsmodels Python package (101). The best fit model was identified based on the change in AICc from the previous model, the Pearson correlation coefficient, and condition number. The age regression was performed in R (102).

Metagenomic prediction was performed using PICRUSt (103). Briefly, the OTU table for a single sample from each healthy adult was filtered to remove any sample less than 1000 reads, and any OTU which recruited less than 1000 reads, or was present in less than 25% of the samples. Metagenomic prediction was performed using PICRUSt KEGG pathways not know to be associated with bacteria were filtered out. KEGGs were collapsed to the level 3 pathways. A Kruskal-Wallis test was used to compare the relative abundance of each KEGG. Maslan analysis was performed using QIIME (78) to filter the 11308 initially observed OTUs to remove singleton and doubleton
OTUs, and OTUs not present in at least 5 samples. We then hierarchically summed the OTUs to allow for testing at multiple taxonomic levels. We used HUMAnN to organize 6906 KEGG orthologs inferred by PICRUSt into 196 KEGG modules and 182 pathways. In order to test the influences of metadata such as diet, sex, and BMI on community composition, we performed multivariate association testing using boosting and a generalized linear model (MaAsLin). We applied MaAsLin to these 6078 filtered clades, 196 modules, and 182 pathways to determine which were significantly associated with 14 curated metadata of interest. A zero-inflated linear model was used for OTUs, and a non-zero-inflated model was used for modules and pathways. Data were filtered to remove individuals who were under three years of age and reported a weight over 39 lbs, a height over 36 inches, or use of alcohol.

Spatial autocorrelation was assessed using Moran’s I statistic (104). Taxa present in less than 10 samples were filtered, since these would not be sufficiently powered. Analyses were conducted using binary spatial weight matrices, with neighborhoods of 0-50 km, 50-100 km, and 100 – 500 km. The different neighborhoods were useful for detecting spatial autocorrelation at difference scales. All spatial weights matrices were row-standardized. We checked for spatial autocorrelation at three taxonomic ranks: class, genus, and OTU. We also considered whether there was autocorrelation within subsets of individuals having IBD, no IBD, diabetes, no diabetes, antibiotics, and no antibiotics. Statistical significance was assessed using permutation tests, which were implemented using a Markov Chain Monte Carlo algorithm. To assess each p-value, 10 chains were run starting each from a different random permutation. Each chain had 100,000 iterations. We used Bonferroni corrections to correct for multiple comparisons, with an overall significance level set to 0.05. Analyses were run using custom Java code, optimized for running many spatial autocorrelation analyses on large data sets (105).

Network Analysis was performed on a single sample for each individual. OTUs were collapsed to the genus level. Genera present in less than 25% of the samples were removed. Spearman correlation of taxon-taxon relative abundance were performed, and only links with an absolute value of correlation greater than 0.4 and an FDR-corrected p value of 0.05 were included. Network
analyses were performed in Cytoscape using an edge weighted spring layout (106). The association of Christensenellaceae with BMI was evaluated using a Mann Whitney U test and chi-squared test.

Statistical power was calculated using the scikit-bio and Statsmodels Python packages (107; 101), using the methodology described in Chapter 2. Samples were paired using amples were matched on age in decade, IBD and diabetes diagnosis, the number of types of plants consumed in a week, categorized BMI, collection season and sleep duration. The samples were subsampled at different depths, and a Kruskal-Wallis or permanova test with 249 permutations were applied to the subsample for alpha and beta diversity, respectively, to calculate the curves. Permutative p values are calculated as $\frac{1 + \text{number more extreme}}{1 + \text{total permutations}}$. The number of permutations was selected to balance the dynamic permutative range and run time. The table was calculated using 99 permutations. Monte Carlo simulations of statistical power were calculated at each sampling depth, where power was defined as the fraction of subsamples that were significant at $p < 0.05$ out of the total number of subsamples. Three rounds of power estimation were performed for each depth. The power estimates and sample sizes were converted to effect sizes using Statsmodels. The average effect size and confidence interval were used to extrapolate power curves.

3.2 The American gut is diverse, and in some cases extreme

The overarching goal of the American Gut Project (AGP) is to characterize the extent of microbial diversity associated with humans, including how this diversity varies with geographic location, health, and lifestyle. The diversity present in a single sample can be described in terms of the abundances of the individual organisms present in the sample. As expected, the body habitat from which the sample is collected has a great impact on the microbial community structure (108). For instance, skin samples typically contain higher abundances of organisms likely to be aerotolerant like those from the Proteobacteria phylum, while fecal samples contain predominantly organisms likely to be anaerobic and are dominated by the Firmicutes and Bacteroidetes phyla. However, the bounds of diversity associated with humans are not known. The best reference microbiome dataset to date was the Human Microbiome Project, which was heavily weighted toward oral and
skin samples from 242 individuals. In contrast, the American Gut is weighted toward fecal samples, collected from nearly 20 times as many individuals encompassing a much wider span of phenotypic diversity (Figure 3.3A, C). The total number of operational taxonomic units (OTUs) observed in the fecal samples is much higher than in the HMP (Figure 3.3B), while the variation in phylum-level abundances for fecal samples is even more extreme than observed in the HMP, with some samples being comprised nearly 100% of Firmicutes, and others being almost entirely devoid of this phylum (Figure 3.3D).

3.3 The Microbiome is influenced by lifestyle decisions and diet

We found correlations between microbial community structure and many demographic, geographic, lifestyle and health-related factors, allowing us to quantify strong versus weak effects on a common scale (Figures 3.4 - 3.7). Fecal microbial communities do not form distinct clusters in an unweighted UniFrac ordination space, a between-sample comparison of community structure based on phylogenetic relationships (beta diversity) (65). This reflects a complex relationship between the host status and their microbial community (Figure 3.4A) and provides no evidence of the distinct “enterotypes” suggested by (109). One of the strongest drivers of the community structure is community alpha diversity (within sample diversity), as measured by Faith’s phylogenetic diversity (PD) (96). No metadata variable alone explains all the variation in alpha and beta diversity (Figure 3.4, 3.7, 3.6; Table 3.4). However, because the sample size is so large relative to previous studies, we can detect more subtle effects on diversity than have been seen previously.

One of the most striking drivers of community structure was the age of the participants (Figures 3.4B, 3.6, 3.5). The gut microbiome is known to be highly plastic during the first two to three years of life, but is relatively stable and resilient once an adult configuration is reached (30; 99; 110). Within the adult American Gut population, we see changes in community structure with age (Figures 3.4B, 3.6A). The microbiome seems to develop with its host long after solid food has been introduced. Interestingly, we find divergent trajectories for men and women (Figure 3.5). While there was no observed difference in alpha diversity between pre- and postmenopausal women
Figure 3.3: The diversity represented by the American Gut Project (AGP). (A) Principal coordinates analysis of unweighted UniFrac distance for the AGP and the Human Microbiome Project (HMP) for fecal (AGP red, HMP Green), oral (AG blue, HMP purple) and skin (AG orange, HMP yellow) samples. (B) Distribution of samples and sequencing depth for the AGP (white) and HMP (grey). (C) Range of microbial beta diversity being observed. Each confidence interval represents the average minimum weighted UniFrac distance between 10 random samples. As the number of samples increases, the chance of observing a similar sample increases leading to smaller intervals. This plateaus to the average minimum distance between all the samples. The number of samples needed to observe a similar sample in the AGP (teal) is much higher that in the HMP (red) indicative of a broader range of microbial diversity being observed. (D) Phylum level distributions for AGP fecal samples. OTUs which did not map to the eight most prevalent phyla were collapsed into the “Other” category.
Figure 3.4: The health and lifestyles of participants in the American Gut Project and their associations with the diversity of gut fecal samples. (A) Fecal samples do not cluster by any metadata variable in unweighted UniFrac PCoA space, although Faith’s Phylogenetic Diversity represents a clear gradient. Darker indicates higher diversity. (B) Age has a significant association with unweighted UniFrac distances within fecal samples. Each bar is the average distance (± stdev) between the reference group, given by the label for the cluster of bars, and the group described by the bar color. Significance was tested with a one-tailed permutation t-test (p < 0.05 *, p < 0.01: **, p < 0.001: ***). (C) Alcohol consumption and phylogenetic diversity are significantly positively correlated (FDR-corrected p < 0.05). (D) In predicted functional profiles from PICRUSt, people who consume less than five types of plants have fecal microbiomes with unique pathways compared to those who consume more plants. Categories shown have an FDR-correct p value less than 0.001. (E) No significant spatial autocorrelation between fecal samples was seen at multiple distance scales.
Figure 3.5: Lifestyle variables modify alpha diversity. A single sample from each individual was used for (A) age, (B) most recent antibiotic use, (C) body mass index. A single sample from healthy adults were compared for (D) Average nightly Sleep Duration, (E) Collection Season, and (F) The number of types of plants consumed in a week. A multivariate regression was use to model alpha diversity, the results of which can be seen in Supplemental Table 3.4.

(age 25 to 45 and 55 to 75, respectively), younger men exhibited an increase in alpha diversity with increasing age from 25 to 45 years, a trend not observed in men in the age range of 55 to 75. This observation underscores the value of the large AGP data, because previous studies with smaller sample sizes have not reported a sex-specific effect on the gut microbiome. However, sex differences have been reported in the composition of the gut microbiota of non-human primates (111; 112; 113). These differences have been hypothesized to be a result of sex hormones and their effects on host physiology (111). Likewise, we hypothesize that the difference observed here between younger and older men may be related to testosterone levels. After the age of 40, men experience a gradual decrease in testosterone levels, which could affect the gut microbiota (114). Interplay between the microbiome and testosterone has been observed in mice. Testicular development is delayed in germ-free mice, and the microbiome suppresses serum testosterone levels in female mice and raises testosterone levels in male mice (115; 116). Additionally, low testosterone levels are correlated with an increased risk for cardiovascular and metabolic disease, and the gut microbiota has been associated with these conditions (117; 15; 118; 119). Thus, it seems plausible that reductions in testosterone in older men may influence the gut microbiota.
The American Gut Project is among the few cross-sectional studies that include samples for individuals with multiple medical conditions, including IBD, diabetes, and obesity (Table 3.2). An effect was observed when comparing lean and obese participants, and Christensenellaceae was associated with lean subjects, as seen in Goodrich *et al* (Figure 3.8) (16). Additional clusters of co-occurring taxons were observed among common Gammaproteobacteria, including Enterobacteriaceae and *Erwinia*, centered around *Pseudomonas*. Another cluster centered around *Peptoniphilus* was also observed, and, as the Christensenellaceae hub, confirms a previous finding (16). The presence of these additional clusters suggests the size of the American Gut cohort facilitates the observation of novel interactions.

Among healthy adults aged 20 to 69 years with BMIs 18.5 to 30 who had not used antibiotics in the past year and reported no diagnosis of IBD or diabetes, lifestyle strongly influenced the microbiome. The effect tended to result from extreme states, especially deficiencies; individuals who reported eating less than five types of plants in a week had lower alpha diversity than their plant-eating counterparts (FDR-corrected Kruskal-Wallis $p < 0.05$) (Figure 3.7, 3.6; Table 3.4). It has been hypothesized that this is due to the number of bacteria associated with plants: a 3-4 leaf spinach plant has more than 850 bacterial OTUs present both on the surface and as endophytes living within the plant (120). PICRUSt metagenomic prediction, which predicts bacterial metagenomes (i.e., presence of genes) from the 16S profile of a sample, showed a difference in the predicted metagenomic profile for the non-plant eaters, with a tendency toward oxidative pathways as has been previously observed with carnivorous mammals (Figures 3.4D, 3.9) (103; 121). In addition, xenobiotic degradation appears to be enriched in non-plant eaters as well (Figure 3.9).

The frequency of alcohol consumption was correlated with increased alpha diversity and changes in community structure, measured by unweighted UniFrac distance (Figures 3.4C, 3.6; Table 3.4) (65). A previous comparison of alcoholics and moderate drinkers found no differences in alpha diversity, yet these differences reflect seemingly contradictory evidence about alcohol consumption: namely that moderate drinking is seen to reduce inflammatory stress and cardiovascular risks, while binge drinking or addictive behavior may exacerbate the same conditions (122; 123; 124).
Figure 3.6: Lifestyle variables modify alpha diversity. A single sample from each individual was used for (A) age, (B) most recent antibiotic use, (C) body mass index. A single sample from healthy adults were compared for (D) Average nightly Sleep Duration, (E) Collection Season, and (F) The number of types of plants consumed in a week. A multivariate regression was used to model alpha diversity, the results of which can be seen in Supplemental Table 3.4.
Figure 3.7: Lifestyle variables have an effect on the microbiome community structure. Each bar is the average distance (± stdev) between the reference group, given by the label for the cluster of bars, and the group described by the bar color. Significance was tested with a one-tailed permutation t-test \((p < 0.05: *, p < 0.01: **, p < 0.001: ***)\) A single sample from each individual was used for (A) body mass index and (B) most recent antibiotic dose. In healthy adults, defined as those 20-69 with a BMI between 18.5 and 30, and no reported history of IBD, diabetes, and no antibiotic use in the past year, (C) alcohol frequency, (D) nightly sleep duration, (E) collection season and (F) the number of types of plants consumed in a week changed the community structure.
Figure 3.8: Co-abundance networks include Christensellaceae as a hub in the healthy subset of adults. Edges represent correlation. Positive correlations are in blue, negative correlations are in grey and the size of the edge is based on the strength of the correlation. The size of the node is based on the number of correlations associated with the corresponding taxon. The presence of Christensellaceae as a hub confirms the same finding in a different cohort (16). We further confirmed the previous finding that Christensellaceae is associated with protection from obesity (16) Mean rank of Christensellaceae relative abundance was significantly lower in lean subjects than in subjects with BMI above 25 (Mann-Whitney U test, p = 0.003), and Christensellaceae was more often present in lean versus overweight subjects (Chi-square test, p = 0.020).
Figure 3.9: Significant associations between KEGG orthologies (KOs) and metadata. Rows correspond to KEGG pathways, and columns correspond to metadata. Heatmap color intensity corresponds to log scaled p-values of associations. Associations marked with (+) are positively correlated with metadata, while those marked with (-) are negatively correlated.
Table 3.4: Regression of PD whole tree diversity against lifestyle variables for adults in the Northern Hemisphere with a BMI of less than 40. The best fit for the model was determined step-wise using the R² value, AICc, and condition number.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>95% CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept*</td>
<td>17.22</td>
<td>[13.38, 21.07]</td>
<td>4×10⁻¹⁸</td>
</tr>
<tr>
<td>Last Antibiotics Past year</td>
<td>-1.93</td>
<td>[-2.83, -1.03]</td>
<td>3×10⁻⁵</td>
</tr>
<tr>
<td>Last Antibiotics Past six months</td>
<td>-1.81</td>
<td>[-2.72, -0.90]</td>
<td>1×10⁻⁴</td>
</tr>
<tr>
<td>Last Antibiotics Past month</td>
<td>-2.92</td>
<td>[-4.37, -1.48]</td>
<td>8×10⁻⁵</td>
</tr>
<tr>
<td>IBD Diagnosis</td>
<td>-4.93</td>
<td>[-6.56, -3.29]</td>
<td>4×10⁻⁹</td>
</tr>
<tr>
<td>Last Antibiotics in IBD Past year</td>
<td>3.01</td>
<td>[-0.66, 6.67]</td>
<td>0.11</td>
</tr>
<tr>
<td>Last Antibiotics in IBD Past six months</td>
<td>1.11</td>
<td>[-2.66, 4.88]</td>
<td>0.56</td>
</tr>
<tr>
<td>Last Antibiotics in IBD Past month</td>
<td>6.55</td>
<td>[2.31, 10.79]</td>
<td>2×10⁻³</td>
</tr>
<tr>
<td>Number of types of plants 6 to 10</td>
<td>1.33</td>
<td>[0.09, 2.56]</td>
<td>0.04</td>
</tr>
<tr>
<td>Number of types of plants 11 to 20</td>
<td>1.91</td>
<td>[0.69, 3.12]</td>
<td>2×10⁻³</td>
</tr>
<tr>
<td>Number of types of plants 21 to 30</td>
<td>1.90</td>
<td>[0.65, 3.16]</td>
<td>3×10⁻³</td>
</tr>
<tr>
<td>Number of types of plants More than 30</td>
<td>2.53</td>
<td>[1.23, 3.83]</td>
<td>1×10⁻¹</td>
</tr>
<tr>
<td>Alcohol Use Less than once a week</td>
<td>1.01</td>
<td>[0.16, 1.18]</td>
<td>0.02</td>
</tr>
<tr>
<td>Alcohol Use Once or twice a week</td>
<td>1.66</td>
<td>[0.76, 2.56]</td>
<td>3×10⁻⁴</td>
</tr>
<tr>
<td>Alcohol Use Three to five times a week</td>
<td>1.37</td>
<td>[0.46, 2.27]</td>
<td>3×10⁻³</td>
</tr>
<tr>
<td>Alcohol Use Daily</td>
<td>1.00</td>
<td>[0.00, 2.01]</td>
<td>0.05</td>
</tr>
<tr>
<td>Exercise Frequency Once or twice a week</td>
<td>1.66</td>
<td>[0.11, 2.21]</td>
<td>0.03</td>
</tr>
<tr>
<td>Exercise Frequency Three to five times a week</td>
<td>1.78</td>
<td>[0.79, 2.76]</td>
<td>4×10⁻⁴</td>
</tr>
<tr>
<td>Exercise Frequency Daily</td>
<td>1.71</td>
<td>[0.65, 2.66]</td>
<td>2×10⁻³</td>
</tr>
<tr>
<td>Exercise Location Outdoors</td>
<td>0.36</td>
<td>[-0.41, 1.12]</td>
<td>0.36</td>
</tr>
<tr>
<td>Exercise Location Depends on the season</td>
<td>-0.95</td>
<td>[-1.89, -0.00]</td>
<td>0.05</td>
</tr>
<tr>
<td>Exercise Location Both</td>
<td>-0.96</td>
<td>[-1.30, 0.60]</td>
<td>0.64</td>
</tr>
<tr>
<td>ln(Age in Years)</td>
<td>2.23</td>
<td>[1.30, 3.17]</td>
<td>3×10⁻⁶</td>
</tr>
<tr>
<td>cos(Collection Month)</td>
<td>-12.04</td>
<td>[-22.24, -1.84]</td>
<td>0.02</td>
</tr>
<tr>
<td>cos(Collection Month) × ln(Latitude)</td>
<td>3.42</td>
<td>[0.64, 6.20]</td>
<td>0.02</td>
</tr>
</tbody>
</table>

The intercept represents an individual who has not used antibiotics in the past year, does not have IBD, eats less than 5 types of plants, does not drink alcohol, exercises less than once a week indoors, sleeps less than 6 hours, and submitted a sample in January.
Alcohol consumption also reflected a general trend seen in lifestyle variables including plant consumption, last antibiotic use, and sleep duration, in which the unweighted UniFrac distance between the samples decreases as alpha diversity increases (Figures 3.7, 3.6). This is not simply a function of sample number: larger groups do not display the convergent behavior with respect to sample similarity. The trend suggests a dysbiosis gradient in which higher diversity leads to a convergent community, while health or lifestyle choices such as those just described lead to a loss of species. This is congruent with a metagenomic model of IBD and obesity, where the loss of peripheral pathways in the microbial community was strongly associated with disease (125).

The American Gut is unique among current microbiome datasets in the geographical distribution of samples: participants live in 49 US states, the District of Columbia and Puerto Rico. However, no spatial autocorrelation or correlation between the microbiome and state of residence or region of the country could be established (Figure 3.4E). Previous studies have shown individuals who cohabitate tend to have more similar microbiome, especially when animals are present in the household (126). This may suggest the gut microbiome is altered through locally, rather than regionally driven mechanisms.

3.4 Lifestyle factors have different impacts on the microbiome

Statistical power measures the probability of finding a significant difference between two sets of observation, given a difference in the underlying populations. A Monte Carlo simulation was used to estimate statistical power for PD whole tree diversity, weighted and unweighted UniFrac distance on the most extreme groups in nine metadata categories in the American Gut population, reflecting health, lifestyle, diet, and demographics (Figure 3.10; Tables 3.5). Body site on the same individual was used a reference, since it is known to have a strong effect on microbiome composition (29).

We observed a profound effect on associated with the diagnosis of inflammatory bowel disease (unweighted UniFrac effect and 95% CI 0.93 [0.86, 0.99]), age (unweighted UniFrac 0.85 [0.79, 0.90]) and the use of antibiotics (unweighted UniFrac 0.86 [0.73, 0.92]). Somewhat more surprising were
the effects associated with diet: plant consumption had a large effect on unweighted UniFrac distance (0.8 [0.7, 0.9]), while a vegetarian or vegan diet did not (0.24 [0.20, 0.28]). Previous work has observed an effect associated with long term dietary patterns (99), although generalized dietary categories, like “vegetarian” are not well correlated with actual food intake (127).

In general, unweighted UniFrac distance was the most powerful metric, followed up PD whole tree diversity, and then weighted UniFrac distance. The few exceptions were observed on factors that had moderate to small effects, such as the frequency of pool and hot tub use. The exceptions were also associated with population p-values greater than 0.05. The finding is somewhat surprising, given that unweighted UniFrac is qualitative while weighted is quantitative. This suggests that species richness has a larger effect in relation to health and lifestyle than species evenness, as previously observed for IBD in (125).

3.5 Using the American Gut as a discovery platform

One attractive component of using crowdsourcing for sample collection via the AG is the ability to assemble large cohorts (or expand upon existing ones) using targeted subproject inclusion. By empowering citizen scientists to participate directly in ongoing experiments – and pay for a portion of experimental costs for sample analysis – the cost and required manpower of assembling the necessary cohorts for microbiome studies is vastly reduced. The AG has collected an unprecedented number of samples using a crowdsourcing model that will continue to be scalable for some time. It is also possible to leverage this infrastructure to collect samples from specialized cohorts by targeted subproject inclusion. The feasibility of this is demonstrated by incorporation of our Intensive Care Unit (ICU), Office Microbial Community Establishment study (OMCE), Autism Spectrum Disorder (ASD), and Mind Microbiome (M&M) pilot programs.

Since the American Gut is a subproject of the EMP (99), all samples were processed using the publicly available and widely used 16S EMP protocol. The intent was to facilitate meta-analyses with minimal technical variability. Any studies that conform to the EMP processing standard can be combined with relative ease for additional insight, or a wider selection of controls.
Figure 3.10: Statistical power estimates in the American Gut Project fecal samples. Power compares the probability for finding a significant ($p < 0.05$) difference between the two most extreme groups for a category compared to the number of samples analyzed in each category. Dotted lines indicate at 95% confidence interval. Body site (oral vs fecal, gray), IBD diagnosis (teal), antibiotic use in the past month compared to not in the last year (brown), non-drinkers compared to daily drinkers (gold), less than six hours of sleep compared to more than eight (brown), lean vs obese BMI (pink), people in their 20s vs 60s (purple), those who consumed less than 5 types of plants compared to thirty or more (maroon), rare vs. daily exercise (orange) and samples collected in winter vs summer (green), were compared. Power was estimated for (A) PD whole tree diversity, (B) Unweighted UniFrac distance and (C) Weighted UniFrac distance over the extremes of ten categories representing health and lifestyle in the AGP. Samples were matched for age by decade, with the exception of babies 0-2, children 3-12, and teens 13-19; IBD diagnosis; diabetes diagnosis; last antibiotic use;
Table 3.5: Power for American Gut Variables. Samples were matched on age in decade, IBD and diabetes diagnosis, the number of types of plants consumed in a week, categorized BMI, collection season and sleep duration.

<table>
<thead>
<tr>
<th>Category</th>
<th>Group 1</th>
<th>Group 2</th>
<th>PD Whole Tree Diversity</th>
<th>Unweighted UniFrac Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>P-value</td>
<td>Effect</td>
</tr>
<tr>
<td>Age</td>
<td>20s (84)</td>
<td>60s (106)</td>
<td>0.03</td>
<td>0.311</td>
</tr>
<tr>
<td>Alcohol Frequency</td>
<td>Never (52)</td>
<td>Daily (56)</td>
<td>0.01</td>
<td>0.47</td>
</tr>
<tr>
<td>Asthma Diagnosis</td>
<td>No (164)</td>
<td>Yes (69)</td>
<td>0.24</td>
<td>NS‡</td>
</tr>
<tr>
<td>Body Mass Index</td>
<td>Lean (257)</td>
<td>Obese (90)</td>
<td>0.002</td>
<td>0.34</td>
</tr>
<tr>
<td>Born by Cesarean</td>
<td>No (166)</td>
<td>Yes (74)</td>
<td>0.94</td>
<td>NS‡</td>
</tr>
<tr>
<td>Cat Ownership</td>
<td>No (333)</td>
<td>Yes (243)</td>
<td>0.76</td>
<td>NS‡</td>
</tr>
<tr>
<td>Chickenpox</td>
<td>No (107)</td>
<td>Yes (229)</td>
<td>0.03</td>
<td>0.153</td>
</tr>
<tr>
<td>Diet Classification</td>
<td>Omnivore (101)</td>
<td>Vegetarian (47)</td>
<td>0.07</td>
<td>0.286</td>
</tr>
<tr>
<td>Dog Ownership</td>
<td>No (347)</td>
<td>Yes (236)</td>
<td>0.72</td>
<td>NS‡</td>
</tr>
<tr>
<td>Dominant Hand</td>
<td>left (79)</td>
<td>right (181)</td>
<td>0.66</td>
<td>NS‡</td>
</tr>
<tr>
<td>Exercise Frequency</td>
<td>&gt; 1× week (46)</td>
<td>Daily (57)</td>
<td>0.18</td>
<td>NS‡</td>
</tr>
<tr>
<td>Exercise Location</td>
<td>Indoors (107)</td>
<td>Outdoors (134)</td>
<td>0.05</td>
<td>0.176</td>
</tr>
<tr>
<td>Flossing Frequency</td>
<td>Never (30)</td>
<td>Daily (57)</td>
<td>0.005</td>
<td>0.61</td>
</tr>
<tr>
<td>Gender</td>
<td>Female (339)</td>
<td>Male (238)</td>
<td>0.59</td>
<td>NS‡</td>
</tr>
<tr>
<td>Gluten Free Diet§</td>
<td>No (285)</td>
<td>Yes (171)</td>
<td>0.85</td>
<td>NS‡</td>
</tr>
<tr>
<td>IBD Diagnosis</td>
<td>No (121)</td>
<td>Yes (43)</td>
<td>2 × 10⁻⁶</td>
<td>0.81</td>
</tr>
<tr>
<td>Lactose Intolerance</td>
<td>No (298)</td>
<td>Yes (149)</td>
<td>0.53</td>
<td>NS‡</td>
</tr>
<tr>
<td>Last Antibiotic Use</td>
<td>Past Month (38)</td>
<td>&gt; 1 year (124)</td>
<td>0.01</td>
<td>0.54</td>
</tr>
<tr>
<td>Migraines</td>
<td>No (177)</td>
<td>Yes (78)</td>
<td>0.25</td>
<td>NS‡</td>
</tr>
<tr>
<td>Multivitamin</td>
<td>No (376)</td>
<td>Yes (250)</td>
<td>0.28</td>
<td>NS‡</td>
</tr>
<tr>
<td>Pool Frequency</td>
<td>Never (111)</td>
<td>Weekly (70)</td>
<td>0.45</td>
<td>NS‡</td>
</tr>
<tr>
<td>Race</td>
<td>Asian (36)</td>
<td>Caucasian (109)</td>
<td>0.004</td>
<td>0.43</td>
</tr>
<tr>
<td>Seasonal Allergies</td>
<td>No (362)</td>
<td>Yes (302)</td>
<td>0.04</td>
<td>0.167</td>
</tr>
<tr>
<td>Season</td>
<td>Winter (220)</td>
<td>Summer (255)</td>
<td>0.09</td>
<td>NS‡</td>
</tr>
<tr>
<td>Sleep Duration</td>
<td>&lt; 6 hours (60)</td>
<td>&gt; 8 hours (69)</td>
<td>0.006</td>
<td>0.47</td>
</tr>
<tr>
<td>Types of Plants§</td>
<td>&lt; 5 (39)</td>
<td>&gt; 30 (84)</td>
<td>0.01</td>
<td>0.295</td>
</tr>
<tr>
<td>Weight Loss</td>
<td>Stable (192)</td>
<td>Decreased (66)</td>
<td>0.61</td>
<td>NS‡</td>
</tr>
<tr>
<td>Weight Gain</td>
<td>Stable (86)</td>
<td>Increased (33)</td>
<td>0.28</td>
<td>NS‡</td>
</tr>
</tbody>
</table>

*aKruskal Wallis p
†Permanova p; 1000 permutations
‡NS: all estimated power values were less than 0.1 and therefore an effect could not be calculated.
§Gluten free diet is not the same as a Celiac diagnosis.
∥Number of types of plants consumed in a week
¶Asian or Pacific Islander
For example, by combining fecal samples from self-reported healthy individuals in the American Gut with fecal samples collected from an Intensive Care Unit (ICU) Microbiome Pilot Project, we were able to observe a marked difference in beta diversity between healthy American Gut participants and ICU patients (Figure 3.11A). This difference is not likely due to run effects, as the ICU and American Gut samples were sequenced on the same sequencing run. These data helped to highlight specific taxa that are differential in ICU patients (regardless of the reason for admission), and include a massive depletion in previously recognized beneficial taxa like *Faecalibacterium* in the ICU patients (Figure 3.11B; FDR corrected Kruskal-Wallis p $3.7 \times 10^{-40}$). Detail on other projects using the American Gut infrastructure can be found in Appendix C.

### 3.6 American Gut as a community process

Through crowdsourcing, the American Gut Project empowers citizen scientists to participate directly in ongoing experiments – and pay for a portion of experimental costs – thereby reducing the cost of sample collection and analysis. It has succeeded collecting more data from the largest funded NIH microbiome effort, and this model has the potential to address scalability issues confronting metagenomic sample collection for other cohorts (29). Because the project is funded by participants, no funding agency mandate to restrict the data analysis to a specific group of investigators applies. Thus the data are released into the International Nucleotide Sequence Database Collaboration (INSDC) as soon as initial quality control steps have been applied, the data analysis performed in a public forum (GitHub), and all code and data are made available in a series of IPython notebooks for reproducible analyses (128). The paper itself was assembled from the Research Network using Google Docs and an open contribution model. Additionally, because the software and data are all available, the ability to make the project a “living analysis” that is updated as new people are added to the project and new tools become available is unique with respect to other projects on this scale.
Figure 3.11: ICU fecal samples compared against fecal samples from self-reported healthy American Gut Project participants. (A) PCoA of unweighted UniFrac distances comparing ICU and American Gut Project healthy fecal samples. (B) Significantly different organisms. A positive value indicates the taxon is significantly depleted (FDR corrected Kruskal-Wallis $p < 1 \times 10^{-10}$; OTUs binned by genus name) in the ICU samples while a negative value indicates the organism is enriched.
3.7 Prospectus

The American Gut Project provides an example of a successful crowd funded project that allows hypothesis generation about the human microbiome on an unprecedented scale, and makes a free resource of thousands of human-derived microbial samples available to any interested researcher (including researchers at companies), educator, clinician, or student. Indeed, several classes at Johns Hopkins, Eureka College, and Hamilton College have already used integration of student-provided samples into the American Gut dataset as a way of teaching about the microbiome. We believe that the community process for acquiring samples, inviting participants to see their own data and participate in the analysis as well as providing samples, and analyzing the results will be adopted by an increasing number of grassroots projects aimed at harnessing the power of citizen science to understand the world around and within our own bodies.

Consequently, we have seen that even self-collected and self-reported information from members of the public can yield interesting association between the microbiome and different conditions, including certain diseases such as diabetes and IBD, and associations with specific medications. In the next chapter, I explore these effects in greater detail in a clinical setting using the example of Parkinson’s Disease. Although neurological conditions such as Parkinson’s might seem surprising to link to gut microbes, there are several studies in progress focused on the gut-brain axis. Even within the American Gut, we saw associations between the microbiome and factors such as the amount of sleep an individual gets. The focus on Parkinson’s, together with a collaboration spanning clinical research centers at multiple geographic sites, allows us to separate the effects of population, Parkinson’s Disease status, and medications using the techniques I introduced in Chapter 2.
Chapter 4

Presence of Parkinson’s disease and therapeutic modalities have independent signatures in the gut microbiome

In the last chapter, I showed how self-collected data in a large population, in conjunction with new statistical techniques, could provide useful associations between the microbiome and various patient attributes. In this chapter, I turn to a specific clinical problem: is the microbiome associated with Parkinson’s disease, and with specific treatments associated with Parkinson’s. This project provided the opportunity to apply these microbiome techniques directly to a clinically relevant condition. This work will be submitted to *Nature Medicine*, where my specific contributions included statistical power calculations, work on the regression models, and writing the paper.

Parkinson’s disease is a neurodegenerative disease of unknown cause, affecting 950 thousand people in the United States alone, and projected to affect 1.34 million by 2050 (129). The human gut microbiome has recently been associated with several other neurodegenerative diseases in association studies in humans, including dementia (130), schizophrenia (131; 132), and autism (133; 134), and, in mouse models, mechanistic work linking the gut microbiome to the development of EAE (135) and autism (136) have uncovered specific pathways.

Two previous small studies looking at Parkinson’s and the microbiome concluded that there were significant differences between Parkinson’s cases and controls (137; 138). However, the purported responsible taxa vary from cohort to cohort, and direct comparisons are difficult due to differences in sequencing technology and analysis, and different subject selection criteria. For example, one study had a preponderance of controls taking Warfarin and statins, which are modulated
by, and therefore may modulate, the microbiome (139; 140). The second had a racial mix, which
can also impact the microbiome, as seen in the American Gut and Human Microbiome Project
(29)(Chapter 3). Although at least one of the studies detected a significant impact on gut micro-
bioime due to of PD medications, the question of whether the case-control effect could be explained
by the treatment effect was left unanswered. As the gut microbiome has been linked to metabolism
of several drugs (19; 141; 140), understanding the effects of treatment on the gut microbiome may
be important for stratifying patients for future therapies.

4.1 Materials and Methods

4.1.1 Patient recruitment and metadata collection

Institutional Review Boards and Human Subject Committees at participating institutions
approved the study. Subjects represented a subset of research participants in the NeuroGenetics
Research Consortium (142), Three of the NGRC-affiliated movement disorder clinics participated
in the microbiome study: University of Washington in Seattle, Emory University in Georgia, and
Albany Medical Center in New York. Subject selection criteria were: having genome wide genotype
and phenotype data, living, willing to provide a stool sample and willing to complete a Gut Micro-
bioime Questionnaire (GMQ). GMQ collected history and current data on factors that could affect
the microbiome. Data on factors that are associated with PD had already been collected. Cases
had diagnosis of PD (143). PD medication data that were used in analysis were those that patients
were on at the time of stool collection. Controls were free of neurological disease. Importantly,
controls were spouses of patients or random convenient samples (caregivers, neighbors); they were
not hospital-based or disease controls who might have had common exposures or medications that
alters the microbiome.
4.1.2 Microbiome Analysis

DNA extraction and sequencing were performed according to the Earth Microbiome Project Protocols, as previously described (86). Operational taxonomic units (OTUs) were picked in QIIME 1.9 using SortMeRNA against the August 2013 release of the Greengenes 16S rRNA gene sequence database (78; 80; 79). Unweighted UniFrac distance on individual groups was compared using a FDR-corrected permanova with 10000 permutations (68). The relative abundance of bacteria were tested using a bonferroni-corrected Kruskal-Wallis test in QIIME (78). Pseudo-effect size and statistical power were calculated using paired Monte Carlo permutations (Chapter 2). Subsampling was based on matching Parkinson’s disease diagnosis, smoking, inflammatory bowel disease diagnosis, sex, diarrhea in the past three months, other neuropathies, and birth by Cesearian. The categories were selected based on significant differences in UniFrac distance.

The multivariate regression (MRM) against UniFrac distance was performed by transforming each of the covariates to a distance matrix with Binary Manhattan Distance. If a pair was discordant for a covariate (i.e. Drug use), the distance was 1, while matched pairs had a distance of 0. The covariates were tested simultaneously under the same model, using a permutative linear regression (144; 145). The significance of covariates was tested with a permutative t test, and over all model p value was calculated with a permutative F test. The reported slopes and intercepts were fitted from the original data.

Logistic regression was performed using the Statsmodels Python Package (101). The OTU table was converted to boolean, filtered to remove any OTU not present in 20% of the samples or any OTU which was present in all Cases or Controls, as this gave an undefined odds ratio. Each OTU was tested using a logistic regression for Parkinson’s disease status and age. Significant OTUs were identified as those with an FDR-corrected $p < 0.05$. 
4.2 Parkinson’s disease alters the microbiome community structure

Using 16S rRNA amplicon sequencing, we were able to show that Parkinson’s Disease has a substantial effect on the gut microbiome. We first tested whether the overall community structure of the microbiome differs between cases and controls with UniFrac, a phylogenetic measure of differences among microbial communities using the dissimilarity of the taxa on a phylogenetic tree, and the total phylogenetic distance covered by a community (PD whole tree) (65; 96). There was no significant difference in PD whole tree diversity between cases and controls, as previously observed with other alpha diversity metrics (139; 140). We found a significant difference in average unweighted UniFrac distance between cases and controls (permutative $p = 0.0001$, 10000 permutations), indicating that the microbial communities are distinct from one another. The standard pseudo effect size for Parkinson’s Disease was $0.68 \ [0.66, 0.70] \ (95\% \ CI)$. This was larger than the effect that we saw with obesity in the American Gut, but smaller than the effect associated with inflammatory bowel disease (Chapter 3; Figure 4.1A).

Examining the specific taxa involved in this overall community difference, we saw a significant decrease in OTUs from genus *Blautia*, *Oscillospira*, unclassified family Lachnospiraceae, and Unclassified Order Clostridiales in cases (FDR-corrected $p < 0.05$). Blautia was also significantly reduced in a logistic regression model, controlling for age (Table 4.1). The loss of *Blautia* in Parkinson’s patients was also identified in a previous study (138). We observed a significant increase (FDR-corrected $p < 0.05$) in OTUs from genus *Prevotella*, the contested family Mogibacterium, genus *Bifidobacterium*, genus *Porphyromonas*, *Anaerococcus*, and *Streptococcus*. The results were confirmed with the logistic regression for all OTUs by *Porphyromonas* (Figure 4.1B, Table 4.1). Previous studies have reported a decrease in *Prevotella* in cases (137; 138), although only a study using a different hypervariable region of the 16S rRNA sequence, (137), and different methods for taxonomic assignment, indicated a significant increase. The abundance of *Prevotella* may be related to confounding factors such as diet (30; 99) or cardiovascular medications. Previous work that identified a significant difference between Parkinson’s cases and controls had substantial differences...
Figure 4.1: Parkinson’s disease has a large effect on the microbiome. (A) Monte Carlo power and effect sizes were calculated for the disease status (purple), location (dark green), and sex (brown) in Parkinson’s disease. Effect sizes were compared to effects in the American Gut (Chapter 3). The effect of Parkinson’s disease on the microbiome is smaller than the effect associated with IBD in the American Gut, but larger than the effect of obesity. (B) The difference in UniFrac distance may be associated with significantly different taxa. Numbers in parentheses are the Operational Taxonomic Unit (OTU) Identifier.
Table 4.1: Odds Ratio (OR) of finding taxa associated with Parkinson’s disease and age.

<table>
<thead>
<tr>
<th>OTU ID</th>
<th>Taxonomy</th>
<th>PD Status</th>
<th>95% CI</th>
<th>Age</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>4297222</td>
<td>g Streptococcus</td>
<td>6.84</td>
<td>[2.71, 17.28]</td>
<td>1.05</td>
<td>[1.01, 1.10]</td>
</tr>
<tr>
<td>3805726</td>
<td>contested family Mogibacterium</td>
<td>4.59</td>
<td>[2.18, 9.65]</td>
<td>0.98</td>
<td>[0.95, 1.01]</td>
</tr>
<tr>
<td>1077373</td>
<td>g Prevotella</td>
<td>3.13</td>
<td>[1.79, 5.48]</td>
<td>0.98</td>
<td>[0.95, 1.01]</td>
</tr>
<tr>
<td>360761</td>
<td>f Ruminococcaceae</td>
<td>2.78</td>
<td>[1.53, 5.04]</td>
<td>1.05</td>
<td>[1.02, 1.09]</td>
</tr>
<tr>
<td>553611</td>
<td>g Bifidobacterium</td>
<td>3.54</td>
<td>[1.86, 6.71]</td>
<td>1.01</td>
<td>[0.99, 1.05]</td>
</tr>
<tr>
<td>495396</td>
<td>g Anaerococcus</td>
<td>3.55</td>
<td>[1.87, 6.74]</td>
<td>1.02</td>
<td>[0.99, 1.05]</td>
</tr>
<tr>
<td>189866</td>
<td>g Blautia</td>
<td>0.31</td>
<td>[0.18, 0.54]</td>
<td>0.98</td>
<td>[0.95, 1.01]</td>
</tr>
</tbody>
</table>

in medications between the two groups: 54% of controls were on statins vs. 20% of cases, and 15% of controls were on Warfarin vs. 1% of cases (137). Given recent implication of the microbiome in cardiovascular disease (119; 146), this confounding factor is of substantial concern.

4.3 The effect of Parkinson’s disease is a combination of the disease itself and treatment

Previous observations show that the microbiome plays a role in drug metabolism (19; 141; 140). In this study, Parkinson’s patients were taking between one and five drugs from seven drug classes. We therefore hypothesized the difference between cases and controls might be a result of treatment, and not the disease. We choose to analyze the influence of treatment on the unweighted UniFrac distance among samples (disregarding the relative abundance information of different taxa, likely to be affected more by treatment than presence of specific microbes). Among the Parkinson’s patients alone, we saw a significant difference in UniFrac distance between individuals taking Catechol-o-methyl transferase (COMT) inhibitors, CoEnzyme Q 10 (CoQ10), amantadine, and anticholinergic (permutative $p < 0.05$, 10000 permutations), demonstrating that individuals on each of these medications came to resemble other patients on the same medication (Table 4.2). The slope associated with anticholinergic and combination of Carbidopa and Levadopa, were negative (i.e. individuals on these drugs were more dissimilar to one another than to randomly chosen individuals), which may simply reflect the small sample size of patients on each treatment.
Table 4.2: A multivariate regression testing the relationship between unweighted UniFrac distance and seven Parkinson’s treatment classes. The model had a p value of 0.0001 and a correlation coefficient of $R^2 = 0.009$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.649</td>
<td>0.212</td>
</tr>
<tr>
<td>Δ(Carbidopa/Levodopa)</td>
<td>-0.0025</td>
<td>0.0005</td>
</tr>
<tr>
<td>Δ(COMT inhibitor)</td>
<td>0.0032</td>
<td>0.00001</td>
</tr>
<tr>
<td>Δ(Dopamine antagonist)</td>
<td>0.00001</td>
<td>0.80</td>
</tr>
<tr>
<td>Δ(MAO B Inhibitor)</td>
<td>-0.00082</td>
<td>0.11</td>
</tr>
<tr>
<td>Δ(Amantadine)</td>
<td>0.0011</td>
<td>0.04</td>
</tr>
<tr>
<td>Δ(Anticholinergic)</td>
<td>-0.0029</td>
<td>0.002</td>
</tr>
<tr>
<td>Δ(CoQ10)</td>
<td>0.004</td>
<td>0.00001</td>
</tr>
</tbody>
</table>

We then performed a regression of factors affecting the gut microbiome in cases and controls, including information about the disease state and treatment (Table 4.3). When controls were introduced into the regression model, controlling for case/control status and treatment, there was a significant difference in community structure associated with case/control status independent of treatment, as well as a difference associated with specific treatments among the cases. Controlling for case-control removed the significance associated with carbidopa/levodopa treatment, but case/control differences for all of the other medications remained significant. These results suggest that Parkinson’s disease has an effect on the microbiome, independent of the effect due to treatment, but also underscore the importance of obtaining microbiome samples on individuals receiving the same medication for different underlying conditions, because medications themselves can have substantial effects (as has for example been demonstrated for proton pump inhibitors (147)).

This study therefore demonstrates for the first time a systematic dysbiosis in Parkinson’s disease patients, independent of Parkinson’s disease treatment (Tables 4.2 - 4.3). Taking treatment into account treatment explicitly in the analysis as a factor is important for several reasons. First, it is more reflective of the multifactorial nature of the microbiome. In work with human patients, it is especially important to use the appropriate statistical tools that allow us to isolate and examine these differences. In this case, we find that the treatment effect can be deconvolved from site effect, and from the overall effect of disease. This result suggests that the effect of Parkinson’s
Table 4.3: Parkinson’s diagnosis and treatment alter microbial community structure. A multivariate regression was used to test the relationship between unweighted UniFrac distance, Parkinson’s disease status and seven Parkinson’s treatment classes. Parkinson’s controls were assumed not to be receiving any drugs. Parkinson’s patients who did not report treatment history were excluded. The model had a p value of $p = 0.0001$ and a correlation coefficient of $R^2 = 0.010$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.647</td>
<td>1.000</td>
</tr>
<tr>
<td>$\Delta$ (Parkinson’s status)</td>
<td>0.002</td>
<td>0.00001</td>
</tr>
<tr>
<td>$\Delta$ (Carbidopa/Levodopa)</td>
<td>-0.0004</td>
<td>0.42</td>
</tr>
<tr>
<td>$\Delta$ (COMT inhibitor)</td>
<td>0.0004</td>
<td>0.00001</td>
</tr>
<tr>
<td>$\Delta$ (Dopamine antagonist)</td>
<td>0.0005</td>
<td>0.08</td>
</tr>
<tr>
<td>$\Delta$ (MAO B Inhibitor)</td>
<td>-0.0007</td>
<td>0.02</td>
</tr>
<tr>
<td>$\Delta$ (Amantadine)</td>
<td>0.002</td>
<td>0.00001</td>
</tr>
<tr>
<td>$\Delta$ (Anticholinergic)</td>
<td>-0.003</td>
<td>0.00001</td>
</tr>
<tr>
<td>$\Delta$ (CoQ10)</td>
<td>0.005</td>
<td>0.00001</td>
</tr>
</tbody>
</table>
per se is distinct from treatment effect, and suggests that pursuing an early microbiome-based biomarker that allows stratification for different treatments of Parkinson’s related treatments would be advisable. These results also suggest that the microbiome field needs to be more cognizant of ways of isolating complex treatment effects. As we show here, collecting information about treatments within a study can help separate drug-related effects from disease-effects.

A current model of Parkinson’s disease assumes that alpha-synuclein aggregation starts in the colon, and travels through the enteric nervous system to the brain. On the basis of our results of the correlations between disease progression and microbiome state, we propose that the Lachnospiraceae produce butyrate, which may modulate alpha-synuclein associated toxicity. The lower relative abundances of *Blautia* and Lachnospiraceae in disease, and the overall loss of Lachnospiraceae as disease progresses may also be a contributing factor in this process. In this context, it is interesting to note that the genomes of members of the family Lachnospiraceae, including *Blautia* and *Coprococcus*, contain genes for fermentation and short chain fatty acid production including butyrate and acetate (148; 149). Short chain fatty acids, including acetate, propionate and butyrate, are able to cross the blood-brain barrier (150), and butyrate acts on peripheral dopamine producing neurons (151).

Several lines of evidence suggest that Butyrate and other histone deacetylase (HDAC) inhibitors are protective in Parkinson’s disease. Butyrate administration in a rotenone-induced *Drosophila* model of Parkinson’s disease decreased early mortality, locomotor impairments, and a dopamine deficiency (152). In a cell culture with transient alpha-synuclein overexpression, toxicity was associated with nuclear translocation. Butyrate and other HDAC inhibitors were protective against alpha-synuclein included apoptosis of dopaminergic neurons, but not affect general apoptosis pathways (153). Alpha synuclein was able to prevent histone acetylation. Consequently, we propose that one mechanism of alpha synuclein toxicity in Parkinson’s disease may be the deacetylation of histones; butyrate and other HDAC inhibitors may temper some of this effect. The decrease or loss of Butyrate-producing bacteria such as *Blautia* in Parkinson’s patients may exacerbate this process, and should be investigated further. However, the present data cannot resolve whether the
loss of *Blautia* is associated with treatment, or with Parkinson’s itself.

We propose future work with a larger cohort, examining both the microbiome and metabolome to untangle the chemical basis underlying some of these taxonomic changes and their effect on gut and brain functions. Both xenobiotic transfers and a defined-community mouse model could be used to examine the role of Lachnospiraceae in Parkinson’s disease, and confirm this hypothesis mechanistically. Nonetheless, the present results are useful for pointing the way towards biomarker development based on the microbiome that could track Parkinson’s’ progression and, potentially, stratify individuals for different treatment modalities.

This chapter showed how the microbiome techniques I introduced, together with other tools from the Knight lab and our many collaborators, can be used to reveal clinically relevant information about the microbiome. A key aspect of this study is linking the microbiome to specific metabolite that particular microbe produces. In the next chapter, I describe the role of microbes in producing a wide range of specialized metabolites, including further exploration of butyrate and short chain fatty acids. I examine the prospects for applying metabolomics research to a wide range of clinical conditions, as a first step toward mechanistic links between the microbiome and metabolome.
Chapter 5

Specialized Metabolites from the Microbiome in Health and Disease


*Co-first authors

In the previous chapter, I showed how the microbiome differs in Parkinson’s disease patients, both as a function of disease and treatment. I proposed a potential mechanism step wherein the loss of butyrate-producing bacteria may modulate dopamine production. Metabolites represent a strong target for mechanistic links between observed microbial species, and health outcomes. In this chapter, I will present techniques to interrogate different bacterial metabolites and explore the role of these metabolites in health outcomes.

In the human body, microbial cells outnumber eukaryotic cells by as many as ten to one and contribute two orders of magnitude more genes to the hologenome (1; 2; 3). The bacterial metagenome contributes to the production of primary metabolites and the conversion of small molecules into secondary metabolites, also called “specialized metabolites”. Products of bacterial metabolism are believed to modulate human health in many ways (20; 154). Crosstalk between microbial and human metabolites influences processes such as nutrient and xenobiotic metabolism (19), protection against pathogens (155), regulation of the enteric nervous system (156), immune regulation (22), resistance against colorectal cancer (157; 158; 159; 160; 161), complex neurological behavior (162), and affect serum lipid and cholesterol levels (163; 164). Metabolic pathways operating in the human body are thus the result of the combined activities of the human genome and
microbial genome.

Diet is important to the microbiota and metabolome, as food is a major source of metabolite precursors. Humanized mice fed a polysaccharide-deficient diet differed in both the microbiome and metabolome from humanized mice on regular chow (165). Conventionally raised mice fed a polysaccharide-deficient diet showed significantly differences in their microbiomes and metabolomes from humanized mice fed the same diet, or conventional mice fed regular (165). In humans, both long term dietary patterns (166; 99) and rapid, extreme dietary changes (167) are reflected in the microbial communities and metagenome. In particular, amino acid intake correlates with changes in the microbiome. For example, increased amino acid intake increased the relative abundance of Bacteroidetes, and changed the metabolome (166; 99). Individuals who ate a primarily animal-based diet with little to no fiber for two days showed decreased levels of acetate and butyrate in the gut compared to those who consumed a plant-based diet over the same time period (167).

Given the way microbes respond to diet, it is unsurprising that both modulate an individual’s energy balance. Kwashiorkor, a severe form of malnutrition, can be transferred from affected people to mice by fecal transplantation (26). Specifically, fecal samples from human twins, where one co-twin was malnourished and the other was not, produced substantially different phenotypes in recipient mice. Animals that received a fecal transplant from the malnourished co-twin lost 30% of their body mass compared to mice with a healthy fecal donor, despite the fact that both groups of mice were fed a traditional Malawian diet. These differences in the microbiome were coupled to differences in amino acid, carbohydrate, and fat metabolism. Dietary responses also differed depending on the microbiome. Ready-to-use Therapeutic Food, a nutrient rich and energy dense dietary supplement, was able to rescue the Kwashiorkor mice from weight loss, but produced little change in the mice receiving the healthy microbiome (26). This study reflects the importance of both microbiome composition and diet in determining the metabolic profile, which, in turn, influences disease outcomes.

At the other metabolic extreme, obesity can also be modulated by the gut microbiome, metabolome, and diet. Ridaura et al (166) transferred fecal material from lean and obese co-
twins into germ-free mice. The metagenomes of lean co-twins were enriched for genes encoding proteins that ferment complex carbohydrates from plant sources. Co-housing lean and obese mice mitigated the obese phenotype in a diet-dependent fashion. A low-fat, high-plant polysaccharide diet promoted weight loss in co-housed obese mice, while a diet high in saturated fat and low in plant materials did not significantly alter the obese phenotype of co-housed animals. Obese animals co-housed with lean mice and fed a low-fat diet also differed in metabolite profiles from controls. The cohoused animals had more short chain fatty acids (SCFAs) including butyrate, propionate and acetate, and produced more bile acids than obese mice housed only with other obese animals. Co-housing with a lean mouse lead to the loss of disaccharides like maltose, malate and maleate, and tyrosine. These alterations suggest the transmission of microbes with the appropriate dietary triggers can lead to changes in microbial and host physiology.

Commensal microbes, present at various mucosal surfaces of the mammalian host, play a critical role in processing environmental signals (e.g., diet, xenobiotics) and relaying “messages” to the epithelium and via the circulatory system. These messages are classically thought of as microbial-associated molecular patterns (MAMPs), common to various microorganisms, or as metabolites involved in paracrine or endocrine signaling to the host. Often, these metabolites can be beneficial. For example, yet unknown secreted factors of commensal skin bacteria can induce production of human antimicrobial peptides such as defensins, which modulate the human innate immune response (19). An important function of the human microbiome is to supply essential water soluble vitamins including vitamin K, vitamin B12, biotin, folate, thiamine, riboflavin and pyridoxine which are then absorbed by the intestines (168; 169). Several recent reports have now identified specific microbial metabolites beyond those traditionally associated with gut bacteria, and which may represent incipient advances in understanding the inextricable chemical link between mammals and their microbiota. Likely, the vast majority of interkingdom molecular communications remain unknown.

Here, we present pathways associated with the production of primary and specialized metabolites, and examine their possible role in pathogenesis. We suggest a framework for studying and
understanding the role of specialized metabolites, focusing on their production in the gut, followed by absorption, then circulation to their target sites (Figure 5.1). Application of this framework would allow better understanding of many pathologies, and potentially reveal novel therapeutic targets.

### 5.1 Metagenomes, metatranscriptomes, and metabolites

The microbial metagenome links taxonomic identification to metabolite production. Horizontal gene transfer between species, and poor species-level resolution in 16S rRNA gene sequencing, can limit the accuracy of direct metabolic predictions (171; 54), although coarse-grained predictions of metabolism from the metagenome has still been useful in both host-associated and environmental contexts (172). Metagenomic analysis catalogs genes within a microbial environment, and helps provide mechanistic explanations for metabolic changes associated with disease. A systems-level approach comparing the metagenomes of individuals with obesity or inflammatory bowel disease (IBD) to healthy controls suggested that the disease states were associated with loss or gain of enzymes which served to initiate or terminate metabolic pathways (125). Furthermore, shifts in the
microbial metagenome translate to changes in metabolic profiles. Mice transplanted with microbes from obese humans had a reduction in butyrate fermentation genes, along with decreased butyrate levels in the cecum (166).

Metagenome sequencing also shows how pathogenic organisms adapt to polymicrobial infections in the host. Sequencing *Rothia mucilaginosa* genomes from cystic fibrosis (CF) patients and healthy controls revealed adaption by encoding multiple genes for lactate dehydrogenase, enabling utilization of lactate produced by the host and co-inhabiting pathogens (173). Other means of adaption included acquisition of genes for phage lysins, modification of genes responsible for horizontal gene transfer, and genes speculated to play roles in modulation of biofilm formation, namely CRISPR elements (173). Metagenomic prediction could potentially be refined through metatranscriptomic and metaproteomic analyses. These methods may provide mechanistic insight into the regulation of enzymes of interest, as the presence of a gene does not guarantee expression of the corresponding protein product. Bacterial protein expression is regulated at the transcriptional and the translational level by many mechanisms (reviewed in (174; 175)). Thus expression-level techniques may be especially useful in examining rapid changes in bacterial function (176; 177; 178). Integrating multi-meta-omic techniques may provide many more direct links between bacteria and their metabolites than are currently known.

5.2 Microbial metabolite synthesis acts in concert with host metabolism

The best-studied microbial pathways influencing human health involve production of short chain fatty acids (SCFAs) including propionate, butyrate, and acetate. The end product of microbial fermentation of complex non-digestible polysaccharide in the colon (179; 22), SCFAs, and their role in metabolism and immunity, have been appreciated for years (see review by (22)). They contribute to protection from infection and inflammation (180; 181; 182), recruitment and maturation of various subsets of immune cells (183; 26; 184), metabolism (185), and the mediation of host-microbe interactions. SCFAs interact with the host in several ways: via specific G-coupled protein receptors (GPR) -41 and -43 (FFAR3 and 2, respectively) (186; 187), inhibiting histone-
deacetylases (HDAC) and thereby altering host gene expression (188), and inducing autophagy (22).

Propionate production is especially important in human health, promoting satiety, preventing liver lipogenesis, lowering cholesterol, and providing anti-carcinogenic activities (189). Pathways leading to bacterial propionate formation in the human gut have been investigated using multiple approaches (PCR amplification by degenerate primers, 16S rRNA gene analysis, tblastn analysis, and microbial physiology). These approaches have highlighted dominant microbes and microbially-influenced propionate pathways in the human gut (190). The succinate pathway may be the most abundant pathway leading to propionate formation in the human gut, as it is present both in Bacteroidetes, the most abundant phylum in the gut, and the less-abundant phylum Negativicutes. The biosynthetic route via the propanediol pathway is found mostly in Lachnospiraceae, and may be the most important pathway utilizing deoxy-sugars such as rhamnose and fucose from host glycans. The acrylate pathway is the least widespread, and Coprococcus catus was the first human gut bacteria in which this pathway was identified. Propionate and butyrate pathways mainly operate in phylogenetically distinct groups of anaerobic bacteria (190). In general, caution should be used in predicting pathways using in silico methods, because the presence of pathways may simply use butyrate or propionate as energy sources rather than produce them (166; 191). Further, the importance of assessing the directionality of pathways along with gene abundance was found to be critical in differentiating between herbivores and carnivores (121). Thus, identifying directionality may also be crucial for understanding the link between microbes and metabolites. These findings highlight that future studies deciphering roles of specific microbial species in human health via biosynthetic pathways from metagenomes should be performed via multiple approaches (121; 190).

Bacterial and host enzymes are linked into complex metabolic pathways. For example, the neurotransmitter serotonin (5-hydroxytryptamine) is a specialized metabolite synthesized through the interaction between the host and its microbiome. Approximately 90% of the serotonin in a human is synthesized in the gastrointestinal tract (192), and serotonin acts locally to regulate gastrointestinal, cardiac, respiratory, and endocrine functions, as well as crossing the blood-brain
barrier (192; 193). At the most basic level, serotonin is derived from tryptophan by tryptophan hydroxylases -1 or -2 in a tetrahydrobiopterin-coupled reaction, followed by a decarboxylation by amino acid decarboxylase (194; 195). Tryptophan is an essential amino acid; it must be obtained from dietary or microbial sources (196). Germ-free (GF) mice have lower levels of tryptophan, serotonin, and indoles than conventionally raised or humanized (i.e., colonized with human microbiota) animals (165). Tryptophan is synthesized from chorismate by members of several bacterial phyla including Proteobacteria, Actinobacteria, and Firmicutes through enzymes whose genes are encoded in complex operons (197). Escherichia coli can synthesize chorismate, a tryptophan precursor, which acts as a branch-point for many microbial metabolic pathways (198; 199). Bifidobacteria are also predicted to have the capacity to synthesize chorismate (198). Consistent with this key role for bacteria, antibiotic treatment altered tryptophan and indole metabolism in rats (200).

Short chain fatty acid and indole production are among the better-characterized microbial metabolic pathways. They also highlight different strategies in metabolic synthesis. SCFAs are fermented from a variety of different sugars via distinct pathways that do not intersect (190). Serotonin and tryptophan biosynthesis are evolutionarily conserved (201), and reflect the interaction of host and microbial symbionts in the production of specialized metabolites. These examples may provide a framework for the characterization of novel metabolic pathways: confirming metagenomic prediction using integrated approaches including microbial genomics, microbial community analyses, and microbial physiology (190).

5.3 Specialized microbial metabolites that mediate disease

The microbiome plays an important role in health and disease (previously reviewed by (202; 203). A shift in balance between a healthy and diseased microbiome leads to pathologic and metabolic conditions including obesity, colon cancer, and IBD. Host-microbial interactions are in part mediated by the immune response (204), and possibly via specialized metabolites (205). The gut microbiota is a virtual endocrine organ (206) that produces myriad secondary metabolites. Some of these metabolites have positive effects on the mammalian host, as in the case of SCFAs
and antibiotics. Although dysbiosis of the host-associated microbiota has been implicated in various conditions (202; 203), the mechanisms by which specific microbes or microbial communities induce disease remain largely unknown. It is possible that specialized metabolites may mediate many physiological states of well being and illness (Figure 5.1).

5.3.1 Microbial metabolites and cardiovascular disease

Gut microbes contribute to atherosclerosis from dietary consumption of red meat, providing a compelling example of microbial metabolism leading to disease in the human host (119; 146). Comparative levels of choline, trimethylamine N-oxide (TMAO), and betaine, three metabolites of dietary phosphatidylcholine, predicted cardiovascular disease risk in a clinical cohort, and promote atherosclerosis in mice. Concentrations of these metabolites also correlated with microbial activity in the gut – administration of antibiotics suppressed their production, and conventionalization with mouse gut microbes rescued the phenotype (119). Trimethylamine (TMA), the precursor of TMAO, is also the product of microbial metabolism of L-carnitine, a compound abundant in red meat (146). Interestingly, vegans and vegetarians consuming L-carnitine produced much less TMAO than omnivores. Specific gut bacterial taxa also correlated with plasma TMAO levels. Finally, supplementing mouse diets with L-carnitine markedly increased plasma levels of TMA and TMAO, and accelerated atherosclerosis. The mechanism by which TMAO accelerates atherosclerosis is unknown, but it may contribute to forward cholesterol transport via upregulation of macrophage scavenger receptors. These studies demonstrate how a metabolite made by the gut bacterial community enters circulation and contributes to disease. Moreover, ‘western’ (high-fat) diets have been associated with intestinal permeability (207), which would facilitate the absorption of microbial metabolites into the circulatory system.

5.3.2 Microbial metabolites and behavior

The gut microbiome-brain axis is an active area of research (208; 208; 208), and has recently received much attention. Observations of the gut microbiome’s influence on brain development and
behavior, which may be transmitted from the enteric nervous system to the central nervous system via the vagus nerve make gut metabolites immediate suspects for mediating gut microbiota-brain interactions (208; 209; 210).

Some bacteria can produce neuroactive metabolites, ranging from serotonin and gamma-aminobutyric acid (GABA), to dopamine and norepinephrine, to acetylcholine and histamine (see (211; 212). GABA, for example, is an inhibitory neurotransmitter that regulates and participates in various functions in the central nervous system in mammals, and is involved in anxiety and depression. Barrett et al (213) reported the production of GABA through metabolism of monosodium glutamate by Bifidobacteria and Lactobacillus spp. isolated from the human gut. However, GABA from the gut cannot cross an unbreached blood-brain barrier (214). Another connection is that probiotic treatment with Lactobacillus rhamnosus increases GABA receptor expression in the hippocampus, and reduces anxiety- and depression-related behaviors in a mouse model (210).

No complete bacterium-metabolite-target pathway has yet been definitively linked to behavioral changes. However, differences in microbial communities are often correlated with changes in the metabolite profile, and perhaps behavior (e.g.; (215; 136)). In a maternal immune activation (MIA) mouse model for autism spectrum disorders (ASD), both fecal bacterial communities and serum metabolomic profiles are different between animals displaying altered versus normal behavior (136). MIA mice also exhibit autistic-like behavioral phenotypes (increased anxiety, increased repetitive behavior, decreased sociability, and decreased vocalization), and increased gut barrier dysfunction. Interestingly, administration of a probiotic, Bacteroides fragilis NTCC 9343, corrected some of the behavioral deficits and restored levels of some serum metabolites. One of these metabolites, namely 4-ethylphenylsulfate (4-EPS), was sufficient to induce anxiety-like behaviors in mice when injected into the circulation, and one probiotic, B. fragilis, which contributes to a gut microbial community that produces less 4-EPS, could also ameliorate some of the negative effects of potentially neurotoxic metabolites.

The beneficial role of SCFAs in gut health has long been appreciated, but they also play a role in host behavior. For example, acetate crosses the blood-brain barrier, where it is taken up
and activates hypothalamic neurons (216). Intraperitoneal administration of acetate suppressed appetite and led to weight loss. However, although butyrate and acetate were mostly beneficial, propionate may be pathogenic. Comparisons between a propionic acid-based animal model of ASD with acquired mitochondrial disorder and a cohort of 213 children with autism specific disorder suggested abnormalities in acyl carnitines via the mitochondrial tricarboxylic acid cycle (217). When injected intraventricularly in rats, propionate induces behavior with certain features of autism (218).

The role of microbial communities in health and disease has been studied extensively in the past decade, but the microbial mechanisms driving many of these processes, and their targets in the mammalian host, are yet unknown. Specialized microbial metabolites originating from the oral, urogenital tract, gastrointestinal tract, and skin microbiota may be the next frontier in elucidating the molecular mechanisms of pathogenesis and symbiosis (Figures 5.1, 5.2).

5.4 Prospectus

With the development of a three-pronged strategy combining metagenomics, metatranscriptomics and metabolomics, we are now entering new era where we can begin to address the physiological role specialized metabolites of the human microbiome in health and disease. Although a small set of such metabolites have been shown to modulate host physiology, the vast majority of such metabolites in humans (suggested to be present by metagenome sequencing) have not been investigated. We propose that the identification of specialized metabolites, their biotransformations by the microbiome and the host, transport, and their origin via metatranscriptomics, and metabolomics in combination with metagenomics, will open new avenues to investigate underlying mechanisms by which the human microbiome influences health and microbial community development. Subsequent in vivo studies may further substantiate the role of these metabolites and provide insights into such mechanisms. Finally, a deeper understanding of the complex chemical crosstalk between the gut microbiota and humans may advance potentially revolutionary therapies for immune, metabolic, and neurologic disorders.
Figure 5.2: Specialized metabolites, synthesized by the gut microbial community from various precursors, reach target sites and mediate health and disease. SCFA: Short-Chain Fatty-Acids; TMAO: trimethylamine N-oxide; DCA: deoxycholic acid; 4EPS: 4-ethylphenyl sulfate
Consequently, in this chapter, I have described some of the range of microbial metabolism and how it links to human disease. Combining the techniques described here, including metabolomics, marker gene survey, metabolomics, metatranscriptomics, and metaproteomics will move conclusions about microbiome research from correlation to mechanism. In the next chapter, I provide examples of how I have applied these principles, together with the techniques I introduced in Chapter 2 and applied in Chapters 3 and 4, to other studies and diseases. Moreover, I provide a paradigm for generalizing the analyses to the large number of diseases in which the microbiome is involved, which were unknown only a decade ago.
Chapter 6

Contributions to Other Studies

In Chapter 4, I discussed a study about the gut microbiome and Parkinson’s disease, a neurodegenerative disorder of unknown cause, which demonstrated differences in the microbiome between Parkinson’s patients and controls, independent of the effects of treatment. I have also worked on several other microbiome studies, focused on disease states. I either contributed to the overall analysis on these studies or I lead small studies that were useful as pilot investigations. Some of these pilot studies were the impetus for the statistical power method presented in Chapter 2, as sample size estimates were needed to expand the studies. However, the results of these studies stand on their own, and help advance our understanding of the role of the microbiome in disease states.

6.1 Ankylosing Spondylitis in Rats

I contributed to a collaboration looking at a rat model of Ankylosing Spondylitis (AS) (published as “HLA-B27 and Human β2-microglobulin affect the gut microbiota of transgenic rats” (219)). AS is an autoimmune arthritis, which affects joints in the spine, hips and shoulders. The condition can spread to peripheral joints, and may also involve inflammation in the eyes, colon and ileum (220). Although AS is a complex polygenetic trait, human leukocyte antigen B27 (HLA B27) is found in approximately 90% of AS patients, compared to 5-6% of the general population (221). The mechanism of action where HLA B27 confers AS susceptibility is not understood, although several hypotheses exist (221; 222). The colitis symptoms and complex etiology lead our
collaborators to consider the role of HLA B27 in shaping the gut microbiome of individuals with Ankylosing Spondylitis, potentially contributing the colitis and arthritis symptoms.

We examined a rat model of AS, using an HLA B27 transgenic Lewis rat (223). These were compared to a β2-microglobulin transgenic and wild type Lewis rat. Animals were raised according to the animal care guidelines at their respective institutions. The animals were housed in multiple fashions, including single housing and cohousing within the same genotype. Rats were sacrificed at 60 to 70 days, and specimens were taken from their cecum lumen and cecum mucosa. The microbiome samples were processed using the EMP protocols (86). OTUs were picked open reference against the August 2013 release of Greengenes using UClust in QIIME 1.8 (80; 78; 81).

Projections of unweighted UniFrac distance in PCoA space showed a significant difference between the transgenic rats (Figure 6.1A, B). This difference appeared to be the result of increase in genus \textit{Paraprevotella} and a decrease in family Rikenellaceae in the HLA B27 transgenic animals. I performed significance testing using a permanova, and identified significant differences between microbial communities. An analysis on a second, larger cohort of rats, raised in the spring, saw a separation between the transgenic and wild type rats, although the clustering pattern was different than the first cohort (Figure 6.1C, D).

The published Lewis rat analysis was part of a larger study, looking at transgenic models in multiple genetic backgrounds. The Fischer rat model of AS separated from the Lewis model along PC1 when samples from the two genotypes were analyzed together in unweighted UniFrac PCoA space. Among the Fischer rats alone, samples clustered based on when they were reared. The first cohort of Fischer rats, raised with the Lewis rats in the main paper in the spring 2013 clustered separately from the second cohort, raised in the late summer/fall 2013 (Figure 6.2). The way the animals were housed, and the intestinal region sampled also had larger influences on clustering in unweighted UniFrac PCoA space than the transgene-driven separation.

One of the major challenges associated with this study is separating technical (i.e. housing) and biological (i.e. transgene-related) effects on the microbiome. Any impact of the HLA B27 transgene on the rats is confounded by the differences associated with the colony and genetic lines.
Figure 6.1: The Transgenic rats rats separate from Wild Type in Unweighted UniFrac PCoA Space. Samples from the (A) cecum lumen and (B) cecum mucosa in the first cohort of Wild type (n = 7, orange), HLA B27/β2 microglobulin transgenic rats (n = 8, blue) and HLA B7/β2 microglobulin transgenic rats (n = 8, red) show a separation in PCoA space based on the genotype. There were significant differences in the cecum lumen between the WT and transgenic animals (p = 0.001), but not between the transgenic strains (p = 0.77) in the cecum lumen. Significant differences were observed between the all three strains in the cecum mucosa (p = 0.001). A second cohort, reared later, also showed separation in PCoA, although the transgenic effect was not replicated in the (C) Cecum Mucosa or (D) Cecum lumen of the second group.
Figure 6.2: The time of rearing drives differences in PCoA space between Fischer Rats. A second strain of HLA-B27 transgenic rats was used during the same study. PCoA project of unweighted UniFrac distance shows a major separation along PC2 between animals reared in the winter (red), and those raised the subsequent summer and fall (blue).
A rat’s microbiota is heavily influenced by their mother, an effect which can be easily confused with genotype in this transgenic model (224). Microbial communities also drift over time, an effect exemplified by differences in the microbiome across rodent facilities (225). Cohousing, or litter mixing can be used to temper the effect of community drift, but are not always sufficient to correct differences (224). This study was useful because it provided an improved understanding of experimental design, including the importance of relative effect sizes across technological and biological effects of interest.

6.2 Short Bowel Syndrome Pilot

Short Bowel syndrome (SBS) occurs when the intestine is injured in a way that prevents adequate digestion and nutritional intake (226). This can lead to long term complications, including vitamin deficiencies, growth delays, anemia, and *Clostridium difficile* (*C. diff*) infection (226). Common causes in young children include intestinal atresia, a birth defect which results in the narrowing or loss of function of a piece or pieces of the intestine; and necrotizing enterocolitis (NEC), an inflammatory condition leading to intestinal necrosis (227; 228). Short Bowel Syndrome is typically treated with nutritional support, although in extreme cases, intestinal transplant may be necessary (226). Transplant currently has a mortality rate of 30% (226).

I lead the analysis on a pilot study looking at Short Bowel Syndrome in children in conjunction with Tracy Grikscheit’s group at Children’s Hospital, Los Angeles. The Grikscheit group is developing an intestinal organoid which can be grown in a donor host, and used to treat SBS. This pilot included approximately 170 physical specimens from humans, zebrafish, and environmental samples. There were 45 human-associated specimens collected from 30 individuals either during the bowel resection, or during a reconnection at least six weeks later. Tissue samples in humans came from several sites along the intestine including the colon, ileum, and jejunum. Specimens were taken either during the original ostomy procedure, or during the reversal more than six weeks later. Specimen were also collected from a zebrafish model of intestinal resection, sham-operated controls and healthy fish (*n*=69) (229). The samples were sequenced using the EMP protocols (86). OTUs
were picked open reference using UClust against the August 2013 release of Greengenes in QIIME 1.8 (80; 78; 81). The structure of the microbiome was evaluated using PD whole tree diversity and unweighted UniFrac distance (230; 231).

One of the major findings of the pilot was a potential difference between the microbiome in children with atresia and NEC. Children with NEC ($n = 15$) tended to have lower alpha diversity than those with atresia ($n = 21$), whether the stoma was placed in the small intestine ($n = 10$) or colon ($n = 5$; Figure 6.3A). A difference in alpha diversity between the two conditions might suggest that, even early in life, the microbiome associated with inflammation is less diverse. Similar findings have been observed with other inflammatory conditions, including inflammatory bowel disease (23).

The small sample size in this study was one motivation for developing a method for power calculations. Although there was no method for determine the number of specimens to compare the effect of Atresia and NEC on alpha diversity, the difference in the mean alpha diversity but non-significant p value suggest a difference between the two conditions, if we had power to detect it. However, at the time of analysis, there was no method for determining how many more specimens would be needed to look at the effect associated with Atresia and NEC.

I also analyzed paired specimens from children ($n = 13$ individuals) and the zebrafish model ($n = 13$ pairs). The paired specimen looked more similar to one another than to samples that were on the connected (“fed”) end of the ostomy or unconnected (“unfed”) end. (Figure 6.3 B,C). Nutrition studies in adults have shown that many short term dietary changes are insufficient to overcome the stable composition of a microbial community (99). These findings underscore the need to separate individual differences in the microbiome from intervention effects due to changing diets. Cross over studies with sufficient washout periods, and detrending individual differences will be important in dietary intervention studies moving forward. In contrast, the American Gut, presented in chapter 3, focused on long term dietary patterns observed in a cross-sectional study.
Figure 6.3: The Cause of SBS changes the microbiome between individuals, although ostomy surgery does not alter individual structure. (A) PD Whole tree alpha diversity was compared for children with small intestine colitis, colonic colitis and colonic atresia. Despite small sizes, the trend suggests that the inflammatory colitis lead to lower alpha diversity that the non-inflammatory atresia. Expanding this pilot could help investigate the differences. Unweighted UniFrac distance between paired samples from (B) humans ($n = 13$) and (C) zebrafish ($n = 13$) were compared with samples at either end of the ostomy (“fed” and “unfed”) using a Kruskal Wallis test. The community structure of paired samples from a single individual were more similar than samples from the fed, or unfed end of the ostomy, suggesting that even in infants, the microbiome retains an individual characteristic.
6.3 Multiple Sclerosis

Multiple Sclerosis causes scarring and demyelination of the central nervous system, leading to neurodegeneration and progressive disability (232). MS typically presents in two disease courses: Relapsing-remitting (RRMS) and Primary Progressive MS (PPMS). RRMS characterized by discrete events, followed by a period of recovery. In contrast, PPMS involves a gradual progression, with no recovery. RRMS is more common, accounting for approximately 85% of MS cases (232). Like many autoimmune diseases, MS is believed to have a microbiome component (233).

We are working on an ongoing consortium study, including groups at the University of California San Francisco (UCSF), Mt Sinai Medical Center, and Caltech. The goal is to determine if any interactions exist between Multiple Sclerosis and the human microbiome. The project utilizes a multi-pronged approach, which includes 16S rRNA sequencing, xenobiotic mouse transplantation, and functional characterization of the immune interactions with microbes isolated for MS patients.

Samples were collected from patients and healthy controls with no history of autoimmune disorders in the MS clinics at UCSF and Mt Sinai. Fecal samples were collected off of toilet paper using dry swabs in a protocol similar to the American Gut Project. The samples were extracted and sequenced using the EMP protocols (86). OTUs were picked closed reference against the August 2013 release of Greengenes using SortMeRNA in QIIME 1.9 (78; 80; 79). PD whole tree alpha diversity and unweighted UniFrac distance were used to compare alpha and beta diversity (230; 231).

Both MS and MS treatment may alter the microbiome in RRMS patients (Figure 6.4). We identified a significant difference between the within-group unweighted UniFrac distance for Copaxane-treated patients (n=34), and the between-group distance between Copaxone treated patients and controls (n=47; p=0.01) and between Copaxone-treated patients and treatment naive patients (n=39; p=0.01; Figure 3). Treatment naive patients were significantly enriched in an OTU mapped to genus Rickettsia. While preliminary, this is a promising finding which suggests that both MS and MS treatment have impacts on the gut microbiome. Effect size estimation, and
Figure 6.4: **Treatment status alters the microbiome in MS patients.** There was a significant difference in Unweighted UniFrac distance between control \( (n = 47) \) and Copaxone-treated \( (n = 34) \) and between Copaxone-treated patients and treatment naive MS patients \( (n = 39) \). Error bars are standard deviation of the distance.

Moving forward, we plan to expand this cohort to improve study power. Collecting information about covariates which influence the microbiome, identified in Chapter 3, and others, such as diet, pet ownership, exercise, and medication history will be important to tease apart individual relationships with MS \( (17; 126; 234) \). We also plan to use time series to look at the way treatment changes the microbiome. Our collaborators at Mt Sinai have started collecting time series from treatment naive patients starting their first therapy. These should help us look for systematic changes in the microbiome across these individuals.

Work on these pilot studies laid an important foundation for the statistical power technique described in Chapter 2. They illustrated the need to be able to predict the number of observations required to observe a significant difference for the effect of interest, especially when working with underpowered data. The importance of careful study design is also highlighted here. Confounding variables with large effects can complicate, or limit biological conclusions due to large differences in the sampled microbiota independent of the effect of interest.

In summary, in this chapter, I briefly describe a series of studies and techniques that can be
applied both to small pilot data sets and larger scale investigations. With these techniques in hand, we are now poised to connect specific changes in the microbiome across many different disorders, especially but not limited to autoimmune disorders, and develop a true understanding of the full scope of the microbiome’s involvement in human health and disease, complemented by mechanistic studies in animal models. In the next chapter, I describe what I have learned from these studies about the current state of the field, how it has developed during my PhD dissertation research, and how it will continue to develop in the future as we apply these techniques more broadly.
Chapter 7

Discussion and Future Directions

The human gut microbiome is an exciting, and rapidly expanding area of research. There has been an exponential growth in the number of citations involving the keyword “microbiome” or “microbiota” in the last decade (235). The first microbiome-based therapy, fecal transplant for recurrent *Clostridium Difficile* infection was approved by the FDA in 2013 (236). This therapeutic is more than 90% effective, and leads to almost immediate resolution of extreme microbial dysbiosis and clinical symptoms (237; 238). In 2014, a predictive model relating the mucosal microbiome to early Crohn’s disease severity was published, representing a useful, large scale application of the microbiome to disease diagnosis (23). The past three years have also seen unexpected links between our microbes and sex-based differences in autoimmunity (115), anorexia (239), the maternal microbiome and autism risk (136), and mental illnesses including depression (240).

These findings represent only the beginning in the discovery of the human microbiome. Substantial work remains to clarify the role of the microbiome in human health. While efforts to standardize and minimize technical effects have begun, the microbiome community has not adopted standard protocols across body sites. Work in larger cohort and population studies will both help untangle complex interactions between medications, lifestyle, disease, and the microbiome and allow the identification of important taxa. Longitudinal studies, which allow an individual to act as their own control, will also help separate these effects. The inclusion of multi-omics technology, like metagenomic profiling of the microbiome, metabolomics, and single cell culture methods provide promise to expand the functional and mechanistic understanding of the relationship between
humans and their bacteria. To aid in this effort, I have developed a new method for estimating statistical power in microbiome studies. The method has been applied to the American Gut Project, the largest open source, cross-sectional microbiome study to date. We not only observed previously undetected lifestyle effects, but we were able to rank them based on effect size. This was made possible by the consistent study design, common processing techniques, and advances in analysis. I also applied the power technique to a study of Parkinson’s disease, to quantify the effect size associated with a new disease. The Parkinson’s study also represents the first time a regression technique was used with UniFrac distance to untangle the role of medication from the role of disease in shaping microbial community structure. Future studies will need to more carefully consider both medication and lifestyle variable to continue untangling the role of individual factors on the microbiome.

The success of identifying lifestyle influences in the American Gut Project is due to the population size. In part, the study’s population size can attributed to the novel approach to funding and participant recruitment. The American Gut is more than an order of magnitude larger than the Human Microbiome Project (29) or MetaHit (241), published microbiome studies of healthy adults. The diversity of the American Gut Population allowed novel lifestyle comparisons, both by the sheer size of the project, and by introducing participants who might otherwise be excluded from traditional case-control microbiome studies. Crowdfunding has allowed the enrollment of anyone interested and able to support the project. However, this funding model comes with its own set of challenges. Meeting the expectations of participants who see themselves as consumers can be challenging from both a data and participant interaction standpoint. Communication and expectation management remain just as critical in crowd-sourced science as they are in traditional grant-funded work. In the future, I predict a successful funding model for scientific research will be mixed. Granting agencies, including the government and private groups will likely continue to fund science in critical areas. However, there is also growing support from the public to take ownership of scientific efforts through crowdfunding and participation in research.

Working in the boundary between grant-funded work and participatory science, I was able to
expand the understanding of lifestyle effects on the human microbiome. I identified novel effects on the microbiome associated with lifestyle factors, like the number of hours a night someone sleeps. The observed effect may be due a circadian effect on the microbiome, suggested by a recent study on the microbiome and jet lag (242). Alternatively, people who experience more stress tend to sleep less, as indicated by a 36 year European study (243).

I was also able to extend the continuum view of how normal behavior moves to aberrant behavior, like in the context of current alcohol use. In the American Gut, we observed a positive correlation between alpha diversity and alcohol consumption, while studies involving alcoholics have observed deleterious effects of alcohol on the microbiome (122; 124). While it is possible some of the daily drinkers had an alcohol use disorder, given that these occur in 13.9% of the population (244), other demographics (Middle aged adults with enough disposable income and interest to participate) may suggest that regular, moderate drinking is likely. This suggests that there may be differential effects of alcohol, depending both on the context in which it is consumed, and on the frequency of consumption. Future research will be needed to address these questions, and untangle the complex pathology of alcohol use and disuse.

Finally, the American Gut provided an opportunity to look at the long-term impact of health history which was previously uninvestigated. The method of delivery plays an important role in the microbiome of infants, the microbiome of babies born by Cesarean resembles the adult skin microbiome, while vaginally delivered babies strongly resemble the vaginal microbiome (245). Since the microbiome of babies develops over the course of their earlier years, and the first microbial contact forms a basis for the microbiome long term, it was believed Cesarean might have long detrimental effects on the microbiota (30; 246). However, we cannot detect an effect of Cesarean on PD whole tree alpha diversity or UniFrac distance in adults. This points toward the need for a large, longitudinal study to comprehensively address the role of Cesarean in shaping the microbiome moving forward.

The novel biological discoveries were made possible through the development and application of new techniques. Microbiome analysis is challenging, given the composition and multivariate
nature of the data. Monte Carlo estimation of statistical power both confirm previous observations, and expand the ability to compare biological factors. These methods can help combat publication bias, where small studies which meet the critical value are published, but studies which do not are considered negative results. It also standardizes the basis of comparison between effects and populations. A single value, the pseudo $d$ ($\hat{\delta}$), can be used across multiple comparisons to give a standardized results. Adoption of statistical power and effect sizes in the microbiome literature could also improve the techniques used in meta analysis, to allow the routine quantification of effects.

Expanding the use of multivariate modeling will also improve the detection of effects on the microbiome. This expands our ability to interrogate complex data and control for influences between effects. A treatment-based analysis would not have been possible in the Parkinson’s’ study, because the combinatorial space associated with the seven treatment types quickly absorbed any statistical power. By instead using regression to analyze the treatment effect, we were not only able to identify changes in the microbiome associated with medication in our moderately sized cohort, we were also able to detect a disease effect. This split between a disease and treatment effect in Parkinson’s disease may also hold true across other studies. Treatment had a large impact on the microbiome in MS patients, and current studies have yet to disentangle these effects from the effect of the disease.

Looking forward, this dissertation lays the groundwork for the ultimate goal of personalized medicine. The findings presented here will need to be validated through rigorous, mechanistic analysis. Technological advancements are required, both in 16S ribosomal sequencing and more advanced techniques. The introduction of qualitative sequencing information will help separate the effect of bacterial abundance from that due to the presence of an organism. Necessary considerations include the 16S ribosomal copy number (247), extraction and primer bias (28; 59). Multi-omic analysis, including metagenomics, metatranscriptomics, metaproteomics and metabolomics will allow more precise identification of bacteria and their functions. Annotation remains a hurdle in all multi omic techniques, so improvements in reference libraries for metagenomes, proteomics and
metabolomics will be crucial to identify candidate bacterial species, genes, proteins and metabolites (248; 249).

Candidate taxa can be grown through advanced microculture techniques. Estimates of taxa which can be identified through culture-based methods in feces vary between 20 and 40%, although more recent studies suggest there may be fewer species (250). In contrast, around 60% (251; 252) of bacterial species have been isolated using microculture techniques. Further development of specialized media will likely increase this percentage. These taxa can be introduced to a mouse model through mono-associated, or in controlled community structures, to examine the effect of individual species on animal outcomes (253). A promising co-culture model from New Zealand may allow higher throughput analysis of candidate bacteria or communities; human epithelial cell cultures can be grown in the presence of anaerobic bacteria (254). Candidate genes can also potentially be validated by cloning them into easy-to-grow species, and using these organisms in culture (254). Additionally, the direct introduction of a metabolite or set of metabolites into a system confirms a critical step in the molecular communication between bacteria and their hosts (255).

In the future, medical decisions will consider multi-omic profiling - including the genome, transcriptome, microbiome, metabolome and cytokine profiles - to tailor medication and dietary regulation of human health. I particularly look forward to work in the area of autoimmune diseases including Rheumatoid Arthritis and Type I diabetes. There is rapidly accumulating evidence for the role of the microbiome in the etiology of these conditions (50; 49; 256; 257; 258). Understanding the role of treatment or a feedback mechanism between physiological states caused by the condition and the microbiome will likely present challenges. For example, deconvolving the effect of high blood sugar on the microbiome from the microbial influences which may help to maintain an immune response to the beta cells of the pancreas in established diabetic patients may be challenging (259). This work may provide rewarding opportunities to both explore new frontiers and make lasting impacts on individuals. However, there is much work left to do before this goal can be achieved. We need to better understand the factors that shape the microbiome, which perturbations have
lasting effects, and how long it takes to see those effects before we can achieve intervention through personalized medicine.
Bibliography


[42] Leach, J. D. Human Food Project: Anthropology of Microbes.


110


Larsen, P. E. et al. Satellite remote sensing data can be used to model marine microbial metabolite turnover. **ISME J** (2014).


Appendix A

Common Abreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGP</td>
<td>American Gut Project</td>
</tr>
<tr>
<td>AS</td>
<td>Ankylosing Spondylitis</td>
</tr>
<tr>
<td>ASD</td>
<td>Autism Spectrum Disorder</td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index</td>
</tr>
<tr>
<td>C. diff</td>
<td>Clostridium difficile</td>
</tr>
<tr>
<td>DCA</td>
<td>deoxycholic acid</td>
</tr>
<tr>
<td>EMP</td>
<td>Earth Microbiome Project</td>
</tr>
<tr>
<td>4-EPS</td>
<td>4-ethylphenylsulfate</td>
</tr>
<tr>
<td>GABA</td>
<td>gamma-aminobutyric acid</td>
</tr>
<tr>
<td>HDAC</td>
<td>Histone Deactylase</td>
</tr>
<tr>
<td>HMP</td>
<td>Human Microbiome Project</td>
</tr>
<tr>
<td>HUMaN</td>
<td>The HMP Unified Metabolic Analysis Network</td>
</tr>
<tr>
<td>ICU</td>
<td>Intensive Care Unit</td>
</tr>
<tr>
<td>IL-1</td>
<td>Interleukin 1; A proinflammatory cytokine</td>
</tr>
<tr>
<td>IL-6</td>
<td>Interleukin 6; A proinflammatory cytokine</td>
</tr>
<tr>
<td>IRB</td>
<td>Institutional Review Board</td>
</tr>
<tr>
<td>KEGG</td>
<td>Kyoto Encyclopedia of Genes and Genomes</td>
</tr>
<tr>
<td>KO</td>
<td>KEGG orthology</td>
</tr>
<tr>
<td>M&amp;M</td>
<td>Mind and Matter Cohort</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>MaAsLiN</td>
<td>Multivariate Association with Linear Models</td>
</tr>
<tr>
<td>MIA</td>
<td>Maternal Immune Activation Model; a model of Autism spectrum disorder</td>
</tr>
<tr>
<td>MRM</td>
<td>Multivariate Regression Model</td>
</tr>
<tr>
<td>MS</td>
<td>Multiple Sclerosis</td>
</tr>
<tr>
<td>NEC</td>
<td>Necrotizing Enterocolitis</td>
</tr>
<tr>
<td>OMCE</td>
<td>Office Community Establishment Study</td>
</tr>
<tr>
<td>PCoA</td>
<td>Principal Coordinates Analysis</td>
</tr>
<tr>
<td>PPMS</td>
<td>Primary Progressive Multiple Sclerosis</td>
</tr>
<tr>
<td>QIIME</td>
<td>Quantitative Insights into Microbial Ecology</td>
</tr>
<tr>
<td>RRMS</td>
<td>Relapsing-Remitting Multiple Sclerosis</td>
</tr>
<tr>
<td>SBS</td>
<td>Short Bowel Syndrome</td>
</tr>
<tr>
<td>SCFA</td>
<td>Short chain fatty acid</td>
</tr>
<tr>
<td>TMA</td>
<td>Trimethylamine</td>
</tr>
<tr>
<td>TMAO</td>
<td>Trimethylamine N-oxide</td>
</tr>
</tbody>
</table>
Appendix B

Algorithms
Algorithm 1: Empirical Power

**Input:**
- `test` a function which takes a matrix of observations and returns a p value
- `samples` a matrix of observations
- `counts` an array containing the number of observations to draw
- `matched` a boolean value stating if the samples are matched
- `ratio` a list of integers which modifies the group size in `counts`
- `alpha` the critical value at which to calculate the power
- `iterations` the number of tests performed to calculate the power
- `runs` the number of times power should be estimated for each depth

**Output:** `power` the probability of getting a p value less than `alpha`

\[\text{power} = \text{an empty matrix with the first dimension the same length as } \text{counts} \text{ and the second dimension of size } \text{runs}\]

\[\text{if } \text{matched} \text{ is True then}\]
\[\text{sub}\_\text{ids} = 1 \rightarrow \text{length } \text{samples}\]
\[\text{else}\]
\[\text{sub}\_\text{ids} = \text{samples}\]

\[\text{for } \text{count in } \text{counts do}\]
\[\text{ratio}_\text{counts} = \text{count} \times \text{ratio}\]

\[\text{for } \text{run in } \text{runs do}\]
\[\text{for } \text{iter in } \text{iterations do}\]
\[\text{sub}\_\text{sample} = \text{choose } \text{ratio}_\text{count} \text{ observations from each group in } \text{sub}\_\text{ids}\]
\[\text{results}[\text{iter}] = \text{Apply } \text{test} \text{ to the } \text{sub}\_\text{sample}\]

\[\text{power}[\text{count}, \text{run}] = (\text{results} < \text{alpha})/\text{iterations}\]

\[\text{return } \text{power}\]
Algorithm 2: Metadata-based sample matching

**Input:**
- *ids*: a list of unique sample ids
- *vars*: a list of unique metadata variables
- *meta*: a matrix of metadata where the value at position \((id, var)\) corresponds to the value for sample \(id\) and metadata variable, *var*
- *category*: the element in *vars* which should be tested
- *order*: a list of the values associated with *category* which should be represented
- *control_cats*: a list of the elements in *vars* used to match the samples
- *sub_size*: an integer number of samples to draw from each grouped category.

**Output:**
- *subsample*: a matrix of sample ids, grouped by their value for *category*, where the first axis is ordered by *order*

1. \(pairs = hash\)
2. \(groups = list\)
3. \(subsample = matrix\) where the first dimension is of length *subsample* and the second is the same length as *order*

**Step One:** Partitioning the metadata

4. for \(id\) in *ids* do
5.     \(id\_group = a\ list\ of\ the\ values\ in\ meta\ for\ id\ and\ control\_cats\)
6.     \(id\_stat = position\ in\ order\ where\ order == meta[id, category]\)
7.     if *id_group* is not a key in *pairs* then
8.         Create a new hash key, where *id_group* is keyed to an empty matrix with a secondary dimension the same length as order.
9.     \(id\_mat = pairs[id\_group]\)
10.    Append *id* to the the *id_stats* column in *id_mat*
11.   for *id_group* in *pairs* do
12.     \(id\_mat = pairs[id\_group]\)
13.     \(num\_pairs = the\ length\ of\ the\ shortest\ column\ in\ id\_mat\)
14.     Append *id_group* to *groups num_pairs* times

**Step Two:** Drawing random samples

15. \(target\_groups = Draw \ sub\_size\ observations\ at\ random\ from\ groups\)
16. for *id_group* in *pairs* do
17.     \(num\_obs = Count\ the\ number\ of\ times\ id\_group\ appears\ in\ target\)
18.     \(id\_mat = pairs[id\_group]\)
19.     for Each vector along the secondary axis of *id_mat* do
20.         \(keep\_ids = Draw\ num\_obs\)
21.         Append the *keep_ids* to the end of the corresponding vector in *subsample*

22. Return *subsample*
Appendix C

American Gut Supplemental Material

C.1 Projects using the American Gut infrastructure

One attractive component of using crowdsourcing for sample collection via the AG is the ability to assemble large cohorts (or expand upon existing ones) using targeted subproject inclusion. By empowering citizen scientists to participate directly in ongoing experiments – and pay for a portion of experimental costs for sample analysis – the cost and required manpower of assembling the necessary cohorts for microbiome studies is vastly reduced. The AG has collected an unprecedented number of samples using a crowdsourcing model that will continue to be scalable for some time. It is also possible to leverage this infrastructure to collect samples from specialized cohorts by targeted subproject inclusion. The feasibility of this is demonstrated by incorporation of our Intensive Care Unit (ICU), Office Microbial Community Establishment study (OMCE), Autism Spectrum Disorder (ASD), and Mind Microbiome (M&M) pilot programs.

C.1.1 Office Microbial Community Establishment

We recently conducted a study at Northern Arizona University where we explored the spatial and temporal aspects of microbial community establishment in the office environment by monitoring bacterial, archaeal, and fungal communities in nine offices in Flagstaff, San Diego, and Toronto. Samples were taken from the offices approximately every other day over the course of a year. Bacterial 16S rRNA and the fungal internal transcribed spacer 1 (ITS1) were amplified and sequenced in order to track the composition and dynamics of these communities in the offices starting from
UV sterilized building materials. Throughout the course of this study, detailed building and environmental parameters were recorded in order to understand the abiotic factors that might drive the composition of microbial communities. However, we suspect that biotic factors, namely the office inhabitants, are also likely driving community establishment. Through the American Gut Project, we tracked the human skin, oral, and fecal microbiome of human inhabitants of our selected offices. The volunteers coordinated directly with the American Gut Project, recording personal data and sample identifiers through the project’s website. The human subject component of this study was completed using the American Gut Project pipeline under the approval of the American Gut Project IRB, and the PIs on the office study will only receive deidentified sample metadata and sequence data for analysis (sequencing is in progress as of this writing). The American Gut Project thus provided the infrastructure to incorporate a human component into this study, which would have been impractical otherwise as it would have been a costly side-endeavor.

C.1.2 ICU Pilot

The gut is hypothesized to play a central role in the progression of critical illness, sepsis and multiple organ dysfunction syndrome (260). Recent clinical evidence indicates alteration of gut microbiome is associated with infectious complications and mortality in ICU patients (261). To date, evaluations of the microbial ecology in the ICU have been restricted to small culture-based studies. These limited studies frequently demonstrate that ICU patients are rapidly colonized with opportunistic pathogens and suffer significant loss of microbial diversity, which has been suggested to be associated with poor outcomes (261). Further, several studies have demonstrated that the effects of commonly used broad-spectrum antibiotics (common in ICU) on the microbiota can be long lasting, with effects lasting weeks to years (262). However the fate of commensal organisms in the ICU, which serve beneficial purposes, is poorly understood. For this reason, a trial with prospective monitoring of the ICU microbiome with comprehensive culture-independent techniques was needed. To address this question, samples were collected from intensive care unit (ICU) patients who were expected to remain in the ICU more than 72 hours at five different intensive
care units (4 Centers in North America). Data was collected on ICU outcomes, antibiotic use, and nutrition intake in conjunction with the Canadian International Critical Care Nutrition Survey (www.criticalcarenutrition.com) and the Translational Pharmacology of Nutrition Program at the University of Colorado School of Medicine. Patients were not excluded from the pilot based on health status. Fecal samples were collected within 72 hours of admission to the ICU, and at ten days or discharge. These samples were run along side American Gut samples in accordance with the methods described in this manuscript. The results of this study begin to address fundamental questions with regards to the impact of critical illness and ICU nutrition delivery on the microbiome. We hope further data on the prognostic role of microbiome alterations on clinical outcome will be generated from these results as analysis continues. This may help guide future interventions via probiotic, prebiotic, fecal transplant therapy and nutrition interventions to reconstitute a normal, potentially beneficial microbiome following critical illness.

C.1.3 Neurological Cohorts: ASD and the M&M

It has long been recognized that cognitive and/or mental dysfunction is common among patients with GI disorders (263; 264). Recent clinical evidence also suggests that modulating GI activity through the microbiome can have an impact on brain functions and behavior in humans and mice (for review see (265; 240; 266; 267)). This is perhaps not surprising given the pivotal role of the gut microbiome in shaping various aspects of mammalian physiology including the immune responses (268) and endocrine/paracrine function (206), which can have direct or indirect impacts on the brain (269; 270). Interestingly, an active neuro-immune response is common in neurodevelopmental and degenerative diseases such as autism spectrum disorders (ASD), bipolar disorder, schizophrenia and Alzheimer’s disease, and might be causally related to disease procession (271; 272; 273; 274; 275; 276; 277; 278). This is particularly relevant because the gut microbiome is fundamental to immune system function, and might underlie, at least partly, the immune activation in neurological diseases. Together, these clinical observations suggest a close relationship between the gut microbiome and neuronal function and mental health diseases.
The recent development of sequencing techniques has allowed massive phylogeny studies of microbiome composition in human fecal and mucoepithelial samples, and provided remarkable insights regarding the dynamic changes of human microbiome. Many of these studies focused on gastrointestinal disease, yet recent work has begun to correlate bacterial signatures in humans with neurological and mental health diseases (for review see (240; 279; 266). However, one major hurdle with many of these studies is the inherent limitation in cohort size, and this might explain the variation seen between studies (for example some studies found high levels of Bacteroidetes to be associated with ASD, while others found high levels of Firmicutes or Prevotella to be correlative).

To overcome the inherent difficulty of assembling participants to study microbial associations along the gut-microbe-brain axis, two groups are using the AG to aid in cohort assembly and sample processing.

**C.1.3.1 ASD Cohort**

The ASD cohort was established to increase the prevalence of ASD patients and matched controls in the AG population. The aim was to recruit 500 ASD patients and ideally 500 matched sibling controls (neurotypical siblings), or parent controls. The aim of this control group was to attain some genetic lineage control, as well as enabling children/adults from the same household to be assessed for microbial similarity. The hypothesis that the microbiome of ASD patients may be influenced by the microbiome of the home environment could therefore be tested. Comparison of matched controls to the rest of the AG population allows for assessment of variables that delineate microbial similarity, e.g. age, diet, geographic location, medical history, etc. This comparison then allows for a more detailed investigation of the ASD-control comparative analysis, so that individual bacterial taxa that are statistical enriched or reduced in the ASD cohort can be co-analyzed against taxa that are statistically enriched or reduced in the neurotypical cohort against the AG population. This study is performed in collaboration with other ongoing ASD population cohorts at the University of Colorado and Stanford, and provides a platform for continued expansion of this important cohort.
C.1.3.2 M&M Cohort

Unlike the ASD and ICU cohorts, that are restricted to participants with ASD or in the ICU, the M&M cohort is optional for any AG participant. The goal of the M&M cohort is to survey the gut microbiome generally for associations with mental health, and is in collaboration with Northeastern University (NU). At NU, enrollment is ongoing in which participants come to the lab to take a broad psychological battery of standardized psychological surveys and donate stool samples, which are collected using the AG kits and processed using the AG sequencing pipeline. To facilitate expansion of the M&M cohort, we have implemented a portal through the AG webpage where donors can choose whether they would like to contribute their sample to the M&M cohort. If so, they are asked to complete the same psychological battery given to participants at NU, enabling data from these participants to be directly compared to that from individuals recruited using more traditional methods. The M&M cohort dataset will be the largest to date surveying human psychology and the gut microbiome (n=500 at NU; with continuous expansion through the AG), and will provide a foundation for other comparative studies of the gut microbiome and human psychology.