

# Power Analysis of Microbiome Data

Michael Agronah

January 28, 2023

## Contents

0.1	Power Analysis of Differential Abundance Studies . . . . .	2
0.1.1	Background . . . . .	2
0.1.2	Proposed research . . . . .	4
0.1.3	Preliminary research . . . . .	4
0.1.4	Challenges . . . . .	8
0.1.5	Project Contributions: . . . . .	8
0.1.6	Conclusion . . . . .	8

->

-> -> ->

## 0.1 Power Analysis of Differential Abundance Studies

### 0.1.1 Background

Power analysis is a very important aspect of planning an effective study design. In microbiome analysis and many control-case studies, an important question is to determine the minimum number of sample sizes needed to detect a desired effect size. Power analysis answers this question. Power is the probability of correctly rejecting the null hypothesis in order to detect an effect size given that it really exist. Several factors influence power. Power depends on the sample size, effect size, the significant or alpha level as well as methodological factors such as experimental design, groups, statistical procedure and model, correlation between time points, response variable, and missing data. In situations where sample size and effect size are known, power analysis could be used to calculate the power of detecting the given effect size.

The goal of differential abundance microbiome analysis is to test the null hypothesis that there are no differences between OTU abundances among groups of interest. The standard procedure in hypothesis testing is to first define a null hypothesis and then choose a significant level. The test rejects the null hypothesis whenever  $p$ -value is below the significant level and fails to reject the null hypothesis whenever  $p$ -value is above the significant level. This technique has two potential drawbacks. First, rejecting the null hypothesis when the  $p$ -value is less than the significant level merely by chance. This results in what is known as the Type 1 error. The second drawback is failing to reject the null hypothesis when the  $p$ -value is greater than the significant level merely by accident. This is also known as Type 2 error. The goal of power analysis is to minimise these errors and to ensure that true effects are detected with high probability.

Differential microbiome analysis involves multiple hypothesis testing. Thus, there is the need for multiple hypothesis correction to correct for the high probability of detecting significant results from testing OTU individually. The standard statistical methods are used to adjust the multiple hypothesis correction (e.g., Benjamini-Hochberg method).

A number of factors poses difficulties when conducting power analysis for differential microbiome studies. First, knowledge of the effect size is required to conduct a power analysis. A typical approach to obtain effect sizes is to use the effect sizes reported in the microbiome literature for power analysis. A problem in microbiome studies however, is that most publications do not report the effect sizes (Kers and Saccenti 2021). Secondly, there is also the speculation of high publication biases in the field of microbiome studies, with work detecting low effects not being published and researches using approaches that may reveal large effects even when other methods show otherwise (Kers and Saccenti 2021). Kers and Saccenti (2021) and Kelly et al. (2015), for instance, have shown that in studies that compare alpha and beta diversities between groups, the choice of the diversity index used determines the effect that will be observed. Consequently, researchers may vie towards choosing diversity measures that yield high effect sizes. These problems makes it difficult to judge the accuracy of those effect size reported in the literature and to use reported effect for performing power analysis.

Thirdly, many differential abundance microbiome studies seen in the literature do not mention conducting power analysis prior to the studies (Kers and Saccenti 2021). All these reasons has led to the believe that microbiome studies are underpowered with exaggerated effect sizes (Kers and Saccenti 2021). This believe may be justified by the lack of reproducibility experienced in the literature of microbiome studies (Kers and Saccenti 2021). Early research on the relationship between the gut microbiota of obese and non-obesity groups for instance, found significant differences in the diversity of the gut microbiota and significant differences in the Bacteroidetes/Firmicutes (B/F) ratio between obese and non-obese people (Ley et al. 2006). Later research conducted using larger samples found only slight differences in microbial diversity and no statistically significant change in the B/F ratio between non-obese and obese individuals (Sze and Schloss 2016). There is therefore a rising need to conduct power analysis in the field of microbiome studies (Kers and Saccenti 2021; Brüssow 2020).

---

In the literature of microbiome studies, attention has been given to power calculations for microbiome studies comparing the diversity measures (for example, alpha, beta and gamma diversities) among groups of interest (Xia et al. 2018). Power calculations for such studies interested in comparing the diversities between 2 groups can be done in R using the **pwr.t.test** function in the pwr package or the **power.t.test** package in basic R. The package **pwr.avova.test** from the pwr package can be used for power analysis for studies involving more than two groups. The functions **power.prop.test** (statmod package), **pwr.chisq.test** (pwr package), **power.fisher.test** (statmod package) and **power.exact.test** (Exact package), can be used for power analysis for studies interested in comparing the abundances of a single taxon across groups.

For studies involving diversity measures for the multivariate cases, Kelly et al. (2015) proposed a framework for PERMANOVA power estimation implemented in the **micropower** package in R and showed that the distance metric selected for a study has a significant impact on the observed effect. Kers and Saccenti (2021) also performed power calculation to demonstrate how different diversity metrics influence needed sample size in power calculation.

Not much however, has been done about power analysis for studies interested in comparing the frequency of all taxa across groups of interest (Xia et al. 2018). A reparameterized Dirichlet multinomial model was developed by La Rosa et al. (2012) to test the hypothesis of mean abundances as well as scales (variance comparison/dispersion) between groups. This paradigm is implemented in the **HMP** package in R. The package's capability includes the capacity to carry out parameter estimates, multiple hypothesis testing, and power and sample size calculations. In this package, the effect size is determined via a modified Cramer's criterion, which is a measure of the distance between the vector mean abundances between groups. The modified Cramer's criterion yields results in the range [0,1]. A criteria value of zero denotes no difference between the groups under study, whereas a value of one denotes the greatest difference possible.

Using simulation studies, La Rosa et al. (2012) demonstrated that power is influenced by effect size and sample size, overdispersion, and, in some cases, sequencing depth. According to the studies, increasing the effect size and sample size increases power while increasing overdispersion decreases power. When effect size, overdispersion, and sample size are held constant in some of the examples in the studies, the number of reads had a positive effect on power.

**HMP** provides point estimates for effect sizes and power and does not compute effect sizes and power for different ranges of abundances. A knowledge of the effect sizes and the power for different ranges of abundances, however, will provide researchers with an insight into the spectrum effect sizes and power to expect in a study design and be able to determine the average effect size and power across control abundances. Further, in a differential abundance study where the effect sizes and power to detect these effects for only specific taxa is of interest, knowing the effects and power for range of abundance will be handy. There is the need to develop a framework that does power and effect sizes calculations for different ranges of abundances.

To the best of our knowledge, there currently exists no R packages that estimate the power and effects for various ranges of abundances. The Negative Binomial Model is commonly used in differential microbiome abundance studies (Xia et al. 2018; Kodikara, Ellul, and Lê Cao 2022), but to the best of our knowledge, there exists no R package for power analysis that is based on this model. In this project, we propose a framework for quantifying effect sizes and power for various control abundance based on the Negative Binomial Model. Two popular R packages that implement the Negative Binomial Model for differential abundance analysis are DESeq2 and edgeR (Xia et al. 2018). For simplicity, this study focuses on the implementation in the Deseq2 package. However, the same analysis can be repeated using the edgeR package as well.

A key feature of our proposed method is that it can handle non-binary covariates. A limitation of both micropower and HMP is that they do not perform power calculation for studies involving non-binary covariates (Kelly et al. 2015). Both egdeR and DESeq2 allow the use use of continuous covariates (Robinson, McCarthy, and Smyth 2010; Love, Huber, and Anders 2014). The methodology proposed in this project takes advantage of the this functionality implemented in egdeR and DESeq2 to calculate the power analysis for different abundance studies involving non-binary covariates. The method proposed in this project will be implemented in R statistical package. Using this packages, researchers can conduct power analysis for differential abundance studies based on the Negative Binomial Model, estimate the number of taxa with high power and calculate the minimal number of sample size required to detect a specified range of effect sizes and power.

### 0.1.2 Proposed research

This research focuses on the power of differential microbiome studies that are based on the negative Binomial model. The specific goals of this project is to

1. estimate effect sizes and power at various OTU abundances in microbiome differential abundance studies that are based on the Negative Binomial Models.
2. estimate the number of OTUs in this studies that have high power
3. estimate sample sizes needed to detect a given effect size and power

### 0.1.3 Preliminary research

This section describes what I have done on this proposed research.

**0.1.3.1 Data collection and processing:** Using the search terms “autism/[All Fields] AND 16S/[All Fields]”, “autism/[All Fields] AND 16S/[All Fields] AND Fecal/[All Fields]”, raw sequence data from 10 projects that examined the microbiome of children with autism spectrum disorder were gathered from the European Nucleotide Archive EBA and the National Center for Biotechnology Information NCBI. The following are the accession numbers for the 10 projects in the NCBI and ENA archives: PRJNA687773, PRJNA624252, PRJNA453621, PRJNA642975, PRJNA355023, PRJNA168470, PRJNA644763, PRJNA578223, PRJNA589343, and PRJEB45948. Children with autism spectrum disorders represent the treatment group in these datasets, whereas children with neurotypical or typical development are the control groups. Adaptor and primer sequences were removed using “cutadapt” function implemented in a bash script. The trimmed sequence from the “cutadapt” were then processed into Amplicon Sequence variants (ASVs) data using the Dada2 pipeline which involves filtering and trimming, error estimation, denoising, merging paired reads and chimeras removal (Chen et al. 2020).

**0.1.3.2 Models description:** The null hypothesis ( $H_0$ ) and the alternative hypothesis ( $H_a$ ) for testing differences between 2 groups (control and treatment) is formulated by  $H_0 : \mu_{control} - \mu_{treatment} = 0$  and  $H_a : \mu_{control} - \mu_{treatment} \neq 0$  respectively, where  $\mu_{control}$  and  $\mu_{treatment}$  denote the means counts in control and treatment groups respectively. Effect sizes, a measure of the difference in the mean can then be tested. The negative binomial distribution has often used to model microbiome count data due to the presence of overdispersion in microbiome count data. Let  $K_{ij}$  denote the count data for the  $i^{th}$  taxa in the  $j^{th}$  sample.  $K_{ij}$  is modeled by negative binomial distribution defined by

$$K_{ij} \sim NB(\text{mean} = \mu_{ij}, \text{dispersion} = \alpha_i), \quad (1)$$

$$\mu_{ij} = g^{-1}(E_{ij}) \quad (2)$$

$$E_{ij} = \sum_r x_{jr} \beta_{ir}, \quad (3)$$

where  $g$  is a link function,  $\beta_{ir}$  are estimates of the effect sizes and  $x_{jr}$  are the covariates. The relationship between the variance of counts and the dispersion is given by  $\text{Var}K_{ij} = \mu + \alpha\mu^2$ . In this project, the estimating procedure implemented in the *DESeq2* package in R is used for estimating  $\beta_{ir}$ ,  $\mu_{ij}$  and  $\alpha_i$ . Details concerning this procedure is stated in the paper by Love, Huber, and Anders (2014) and by Anders and Huber (2010).

**0.1.3.3 Modelling the relationship between control abundances and fold changes:** Plots of the control abundances and effect sizes are shown in figure 2. Variations around the smooth curve are greatest in the middle and lowest at the ends of the smooth curves. A scale-location plot shown in 2 shows that variation exhibits as quadratic behaviour, confirming the relationship seen in the plot of control abundances fold changes 2. Thus, a quadratic function be used to describe the variance of the effect sizes as a function of control abundances. The variation in effect sizes is models by the scale parameter of a Cauchy distribution. The scale parameter is defined as a quadratic functions of the control abundances, shown in equation by 4.

$$y_i = \text{Cauchy}(\text{location} = 0, \text{scale} = \gamma_i), \gamma_i = \exp(k_0 + k_1 x_i + k_2 x_i^2), \quad (4)$$

where  $y_i$  and  $x_i$  are the effect size and control abundance, respectively, for the  $i^{th}$  taxa and  $k_j; j = 1, 2, 3$  are constants to be estimated.

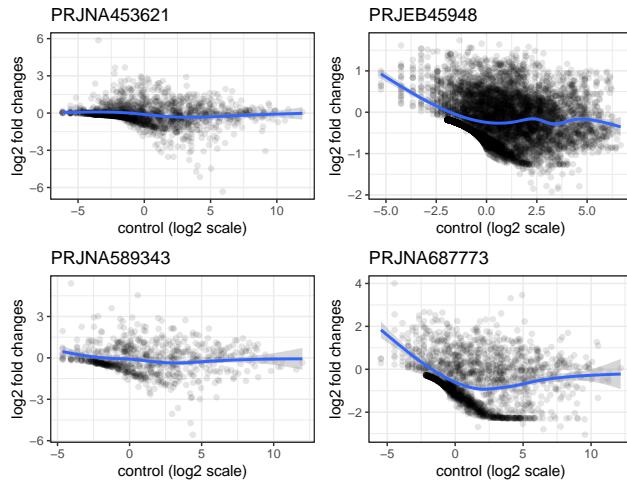


Figure 1: Plot of control against effect sizes for datasets with assessment numbers PRJNA453621, PRJEB45948, PRJNA589343 and PRJNA687773

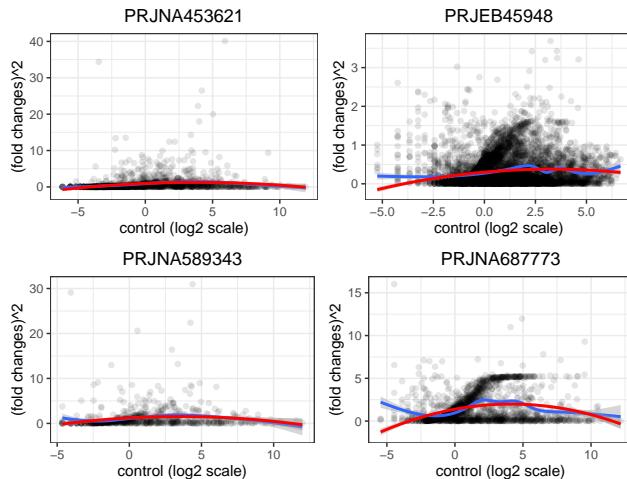


Figure 2: Scale-location plot of control against squared fold changes for datasets with assessment numbers PRJNA453621, PRJEB45948, PRJNA589343 and PRJNA687773

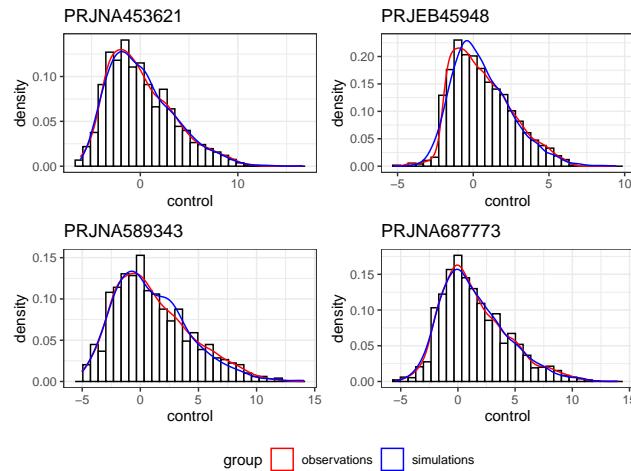


Figure 3: Histogram and density plots for control abundances and density from a simulated sample from fitting a truncated normal distribution

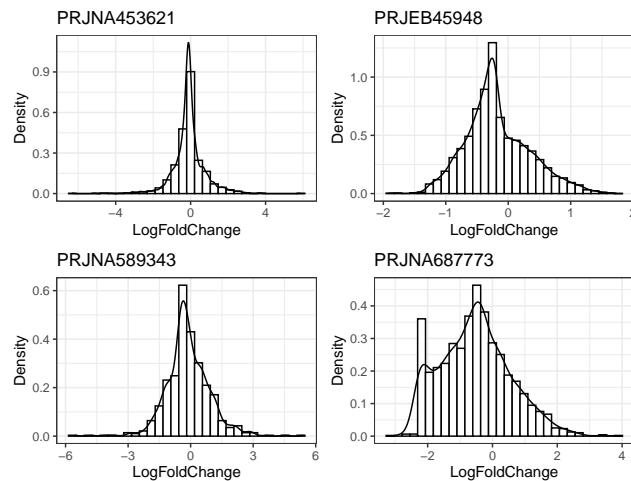


Figure 4: Histogram and density plots for effect sizes

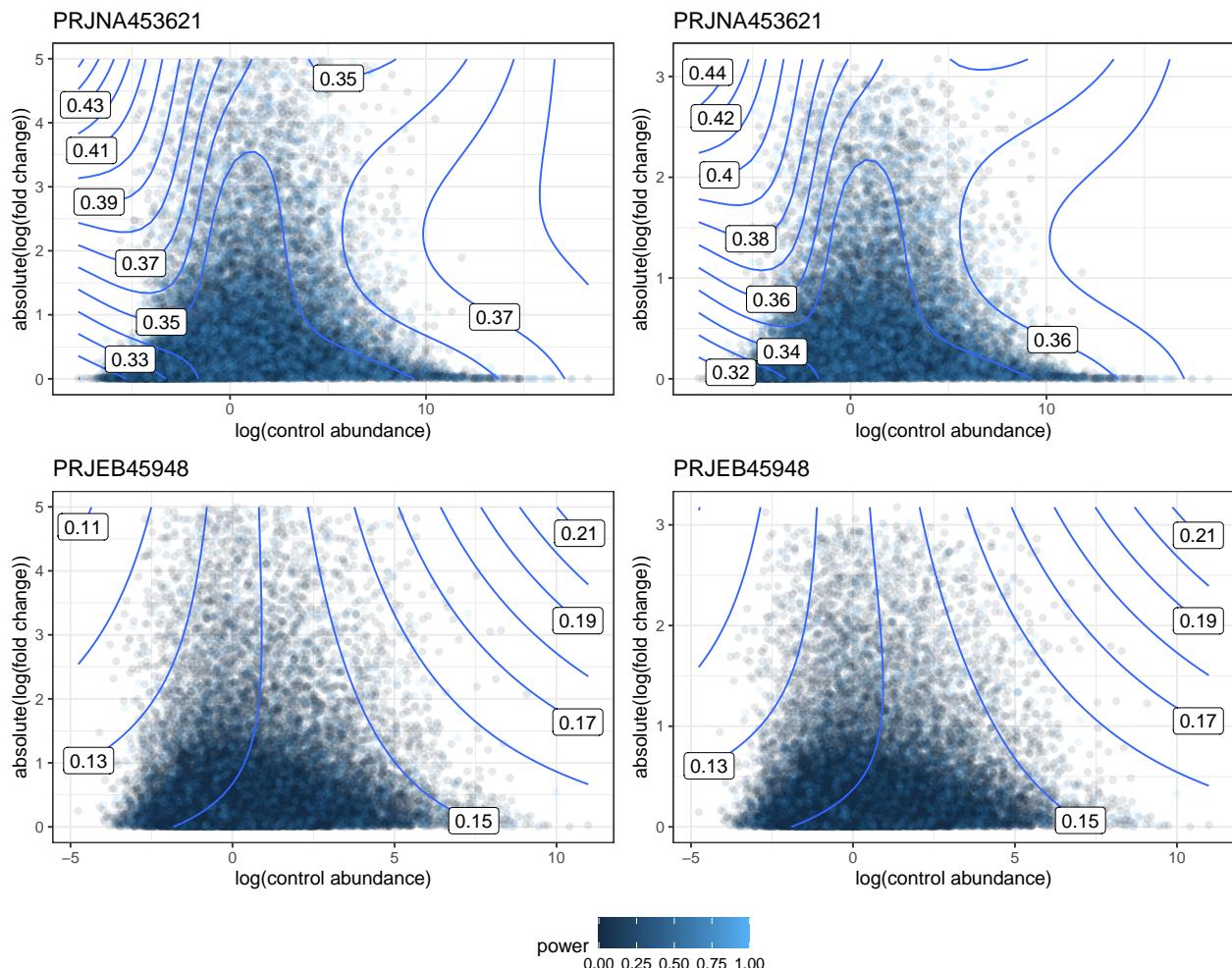


Figure 5: Contour plot showing control abundances and effect sizes with power as contours

#### 0.1.4 Challenges

To perform power analysis, one must first determine the effect size and sample size. A common approach is to obtain the effect size reported in the literature to conduct power analysis and sample size calculation. Most differential abundance microbiome studies in the literature, however, do not report the effect sizes (Kers and Saccenti 2021). There is also the speculation of publication biases with many microbiome studies reporting only results that produce high effects even with low sample sizes (Kers and Saccenti 2021). Further, microbiome studies are known to be underpowered (Kers and Saccenti 2021; Yap et al. 2021). Additionally, reported effects in the literature do not report effect sizes and power for different OTU abundances. Knowing the effect sizes for different OTU abundances would help researchers to estimate the range of effects and power across various OTU abundances and not lump all OTUs as having the same effect sizes or power.

There is therefore the need for a framework that estimates effect sizes for differential abundance microbiome studies for performing power analysis and sample size calculation.

This project proposes a framework for estimating the effect sizes and power for different ranges of OTU abundances in a differential abundance microbiome studies. The method proposed in this project is based on the Negative Binomial Model. To the best of our knowledge, there exists no current work or R package that estimates power and effect sizes for various OTU abundances in differential abundance studies.

#### 0.1.5 Project Contributions:

This project advances the field of differential abundance microbiome research. Based on the Negative Binomial model, we propose a method for estimating the effect sizes and power of individual OTU abundances. The method described here will be implemented in R statistical packages. Using this framework, researchers will be able to tell the spectrum of effect sizes and power for different OTU abundances, gaining insight into how effect sizes and power are distributed across different OTU abundances. In applications where researchers are interested in differential abundance of specific OTUs, for instance, they will be able to estimate the effect size and power of the desired OTUs using the proposed framework in this project. Researchers will also be able to predict the number OTUs that have high power.

A special feature of the method proposed in this project is that it allows for the use of non-binary covariates. This approach, for instance, can be used to determine how a continuous covariate affects the differential abundance of OTUs. Two widely used R packages for power calculation; HMP and Micropower, do not support non-binary covariates. Our approach makes use of the DEseq2 framework, which supports non-binary covariate, to calculate power and effect size.

#### 0.1.6 Conclusion

-> -> ->

- Anders, Simon, and Wolfgang Huber. 2010. “Differential Expression Analysis for Sequence Count Data.” *Nature Precedings*, 1–1.
- Brüssow, Harald. 2020. “Problems with the Concept of Gut Microbiota Dysbiosis.” *Microbial Biotechnology* 13 (2): 423–34.
- Chen, Yu, Hui Fang, Chunyan Li, Guojun Wu, Ting Xu, Xin Yang, Liping Zhao, Xiaoyan Ke, and Chenhong Zhang. 2020. “Gut Bacteria Shared by Children and Their Mothers Associate with Developmental Level and Social Deficits in Autism Spectrum Disorder.” *Mosphere* 5 (6): e01044–20.
- Kelly, Brendan J, Robert Gross, Kyle Bittinger, Scott Sherrill-Mix, James D Lewis, Ronald G Collman, Frederic D Bushman, and Hongzhe Li. 2015. “Power and Sample-Size Estimation for Microbiome Studies Using Pairwise Distances and PERMANOVA.” *Bioinformatics* 31 (15): 2461–68.
- Kers, Jannigje Gerdien, and Edoardo Saccenti. 2021. “The Power of Microbiome Studies: Some Considerations on Which Alpha and Beta Metrics to Use and How to Report Results.” *Frontiers in Microbiology* 12.
- Kodikara, Saritha, Susan Ellul, and Kim-Anh Lê Cao. 2022. “Statistical Challenges in Longitudinal Microbiome Data Analysis.” *Briefings in Bioinformatics* 23 (4): bbac273.
- La Rosa, Patricio S, J Paul Brooks, Elena Deych, Edward L Boone, David J Edwards, Qin Wang, Erica Sodergren, George Weinstock, and William D Shannon. 2012. “Hypothesis Testing and Power Calculations for Taxonomic-Based Human Microbiome Data.” *PloS One* 7 (12): e52078.
- Ley, Ruth E, Peter J Turnbaugh, Samuel Klein, and Jeffrey I Gordon. 2006. “Human Gut Microbes Associated with Obesity.” *Nature* 444 (7122): 1022–23.
- Love, Michael I, Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 1–21.
- Robinson, Mark D, Davis J McCarthy, and Gordon K Smyth. 2010. “edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data.” *Bioinformatics* 26 (1): 139–40.
- Sze, Marc A, and Patrick D Schloss. 2016. “Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome.” *MBio* 7 (4): e01018–16.
- Xia, Yinglin, Jun Sun, Ding-Geng Chen, et al. 2018. *Statistical Analysis of Microbiome Data with r*. Vol. 847. Springer.
- Yap, Chloe X, Anjali K Henders, Gail A Alvares, David LA Wood, Lutz Krause, Gene W Tyson, Restuadi Restuadi, et al. 2021. “Autism-Related Dietary Preferences Mediate Autism-Gut Microbiome Associations.” *Cell* 184 (24): 5916–31.