BRIGHAM AND WOMEN'S HOSPITAL
HARVARD MEDICAL SCHOOL

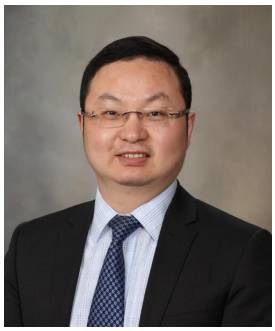
181 Longwood Avenue
Boston, Massachusetts 02115-5804

Department of Medicine
Channing Division of Network Medicine

Channing Microbiome Seminar

June 4th (Friday), 2021, 9AM (ET)

Zoom link: https://us02web.zoom.us/j/86396416935?pwd=d0REZytBZzBvcFhkOHNYOW9selJBdz09
Meeting ID: 863 9641 6935
Passcode: 711645



Jun Chen, PhD

Department of Quantitative Health Sciences
Microbiome Program
Mayo Clinic

Linear Models for Differential Abundance Analysis of Microbiome Compositional Data

Abstract: Differential abundance analysis, which aims to identify microbial taxa whose abundance covaries with a variable of interest， is at the center of statistical analyses of microbiome data. Although the main interest is in drawing inferences on the absolute abundance, i.e., the number of microbial cells per unit area/volume at the ecological site such as the human gut, the data from a sequencing experiment reflects only the taxa relative abundance in a sample. Thus, microbiome data are compositional in nature. Analysis of such compositional data is challenging since the change in the absolute abundance of one taxon will lead to changes in the relative abundances of other taxa, making false positive control difficult. Here we present a simple, yet robust and highly scalable approach to tackle the compositional effects in differential abundance analysis. The method only requires the application of established statistical tools. It fits linear regression models on the centered log-ratio transformed data, identifies a bias term due to the transformation and compositional effect, and corrects the bias using the mode of the regression coefficients. Due to the algorithmic simplicity, our method is 100-1000 times faster than the state-of-the-art method ANCOM-BC. Under mild assumptions, we prove its asymptotic FDR control property, making it the first differential abundance method that enjoys a theoretical FDR control guarantee. The proposed method is very flexible and can be extended to mixed-effect models for the analysis of correlated microbiome data. Using comprehensive simulations and real data applications, we demonstrate that our method has overall the best performance in terms of FDR control and power among the competitors. We implemented the proposed method in the R package LinDA (https://github.com/zhouhj1994/LinDA).

Bio: Dr. Chen is an Associate Professor of Biostatistics in the Department of Quantitative Health Sciences and the lead statistician in the Microbiome Program at Mayo Clinic. He received his PhD in Genomics and Computational Biology from the University of Pennsylvania and did his postdoctoral training in biostatistics at the Harvard T.H. Chan School of Public health. Dr. Chen's current research focuses on the development and application of powerful and robust statistical methods for genomic data analysis. He is particularly interested in methodological development for microbiome data. He has developed many quantitative tools for analysis of high-dimensional structured data including methods for distance-based testing (GUniFrac), kernel machine association tests (SSKAT), structure-adaptive multiple testing (CAMT) and confounder adjustment in multiple testing (2dFDR).

Hosted by Jacqueline Starr