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

Sep 2, 2016, 11am @ 5th floor conference room

Joseph N Paulson, Ph.D.
Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute
Department of Biostatistics, Harvard TH Chan School of Public Health

Accounting for artifacts in high-throughput sequencing data

We introduce a methodology to assess differential abundance in sparse microbial marker-gene survey data. Our approach, implemented in the metagenomeSeq Bioconductor package (bioconductor.org/packages/metagenomeSeq/), relies on a novel normalization technique and a statistical model that accounts for undersampling—a common feature of large-scale marker-gene studies. Using simulated data and several published microbiota data sets, we show that metagenomeSeq outperforms the tools currently used in this field. We motivate this methodology using a large diarrheal cohort. We then show how sparsity impacts other commonly applied methods like PCA, alpha-diversity, etc. We point out several innate characteristics of what the abundance data look and how other types of data - including large heterogeneous RNA-Seq data - behaves similarly and our approach (https://github.com/quackenbushlab/yarn).

Bio: I am a Research Fellow in the Department of Biostatistics and Computational Biology at the Dana-Farber Cancer Institute and Department of Biostatistics at the Harvard TH Chan School of Public Health under the guidance of Professor John Quackenbush. Prior to joining the Quackenbush lab I was a National Science Foundation Graduate Research Fellow at the University of Maryland, College Park where I received my Ph.D. in Applied Mathematics, Statistics and Scientific Computation under the guidance of Mihai Pop and Héctor Corrada-Bravo.

Hosted by Yang-Yu Liu