Evaluation of Statistical Methods for Differential Expression Analysis in Microbiome Metatranscriptomics Data

Di Wu

University of North Carolina at Chapel Hill
E-mail: dwu@unc.edu

Abstract: Differential abundance (DA) analysis in metagenomics (from whole genome shortgun DNAseq of bacteria) and differential expression (DE) in metatranscriptomics (from RNAseq of bacteria) are critical means for understanding differences between micorbiome sample groups (e.,g disease vs normal). However, modeling microbiome data is challenging because of their sparsity (i.e., zero inflation), over-dispersion, high dimensionality, and their inherent hierarchical compositional structure.

Although various statistical testing methods have been used in DA analysis at the taxa level of metagenomics data, how their performance in DE analysis of metatranscriptomics remains unclear. Therefore a comprehensive evaluation of these methods is required regarding metatranscriptomics data at the bacteria gene level may be more zero-inflated and dispersed. We here symmetrically evaluated five methods including logistic beta of compositional data and a few rank-based methods, using simulations based on the real-data inspired zero inflated negative binomial distribution, with different scenarios in terms of mean shift, dispersion, zero percentage and batch effects. Both Type 1 error rate and power are considered. We will use the evaluation results to guide the usage of methods for metatranscriptomics DE analysis.