



Department of Biostatistics

Bernard G. Greenberg Distinguished Lecture Series

Presented by

Robert John Tibshirani
Professor of Public Health Sciences and Department of
Statistics
Stanford University

Finding consistent patterns: A nonparametric approach for identifying differential expression in RNA-Seq data

ABSTRACT

I discuss the identification of features that are associated with an outcome in RNA-Sequencing (RNA-Seq) and other sequencing-based comparative genomic experiments.

RNA-Seq data takes the form of counts, so models based on the normal distribution are generally unsuitable. The problem is especially challenging because different sequencing experiments may generate quite different total numbers of reads, or “sequencing depths”. Existing methods for this problem are based on Poisson or negative-binomial models: they are useful but can be heavily influenced by outliers in the data.

I introduce a simple, non-parametric method with resampling to account for the different sequencing depths. The new method is more robust than parametric methods. It can be applied to data with quantitative, survival, two-class, or multiple-class outcomes. I compare our proposed method to Poisson and negative-binomial based methods in simulated and real data sets, and find that our method discovers more consistent patterns than competing methods. This is joint work with my graduate student Jun Li.

Place: Blue Cross Blue Shield Auditorium

Michael Hooker Research Center

Date: Wednesday, June 6, 2012

Time: 3:30 – 4:30 PM