

COLLOQUIUM

Time: Friday, February 3, 2017 (4:00 PM)

Location: Parker Hall 249

Prioritizing Hypothesis Tests for High Throughput Data

Dr. Sangjin Kim

Department of Biostatistics and Bioinformatics
Duke University

Abstract

The massive increase in the resolution of biological data comes at the cost of increased multiple testing correction. As a result, the power for individual hypothesis tests often becomes very low. One solution is to use independent sources of information to identify the features of the data most likely to be true effects. Such filtering techniques have been in common use throughout the genomic era, but little is known about their statistical properties or the conditions that make them successful. We introduce a framework for quantifying the effectiveness of filtering strategies and show that a filter with a high probability of ranking true positives features highly (e.g. top 10%) can increase discovery probabilities for low effect size features dramatically, but filter effectiveness is highly dependent on the choice of filter cutoff point. We introduce a data-based method for choosing the filter cutoff point in a statistically rigorous fashion and show that its use can increase discovery probabilities by several-fold relative to a naïve choice of filter cutoff. We show various other results relating the filter properties to its effectiveness in increasing discovery probabilities.