Nonparametric, Hypothesis-Based Analysis of Molecular Heterogeneity for Comparative Phenotype Characterization

Jeanne Kowalski, Ph.D.
John Hopkins University

Abstract:

Advances in technology have led to an explosion of molecular research in many fields. Oncology researchers study molecular markers for diagnostic tools by relating expressions from thousands of genes to cancer status, while HIV researchers study drug resistance by relating genetic mutations to altered drug susceptibility. Both tasks include statistical issues of high dimensionality coupled with small sample sizes and thus preclude formal hypothesis testing based on conventional principles.

In this talk, I describe two novel, inference-based approaches to analysis of molecular heterogeneity associated with phenotypes. A common theme among them is the construction of testable hypotheses with assumptions that reflect the complex structure of genetic data. With a modest sample, I discuss a distance-based approach to analysis of genetic heterogeneity based on population sequence data. With the extreme case of several single samples that are to be compared from a microarray experiment, I introduce a stochastic linear hypothesis approach to estimate a number of genes that meet several criteria, beyond experimental variation. In each setting, I also discuss bioinformatics approaches to characterize genes or locations and mutation patterns that depict phenotypes. As motivation for the methods, I examine two separate problems, one for relating differences in a region of the HIV genome to drug resistance, and a second for relating gene expressions with hypothesized pathways for immunogenetic analysis of T cells.