Statistical Challenges in Comparative Proteomics

Marc Coram*, Pei Wang, Hua Tang
* Department of Health Research and Policy, Stanford University

Abstract

Proteomics research has an enormous clinical and scientific potential. However, statistical challenges must be overcome before the full potential can be realized. In this talk, we analyze mass spectrometry data on human blood serum. The samples are chemically treated, separated by liquid chromatography, and analyzed via electrospray ionization mass spectroscopy (LC/ESI-MS). We will describe the statistical methodologies we have developed to analyze these data. Ma jor challenges include: (1) detecting peptide signatures, (2) quantifying peptide intensity, (3) comparing proteomic profiles from run to run. These challenges require scanning each vast and noisy image for the chains of peaks that mark peptide locations. There are several complications. First, two types of noise are present: instrumental noise and chemical noise. Second, sample preparation can substantially distort the signal and noise intensities from run to run. Third, while the mass/charge aspect of the data is quite accurate and reproducible, the chromatographic aspect varies both globally, across each run, and locally, from peptide to peptide. We describe some current approaches to these challenges, and our work on a more unified approach.