Bayesian Methods for Proteomic Biomarker Discovery using Functional Mixed Models

Jeffrey S. Morris

The University of Texas MD Anderson Cancer Center

Various proteomic assays yield spiky functional data, for example MALDI-TOF and SELDI-TOF yield one-dimensional spectra with many peaks, and 2D gel electrophoresis and LC-MS yield two-dimensional images with spots that correspond to peptides present in the sample. In this talk, I will discuss how to identify candidate biomarkers for various types of proteomic data using methods based on the Bayesian wavelet-based functional mixed models. This approach models the functions in their entirety, so avoid reliance on peak or spot detection methods. The flexibility of this framework in modeling nonparametric fixed and random effect functions enables it to model the effects of multiple factors simultaneously, allowing one to perform inference on multiple factors of interest using the same model fit, while adjusting for clinical for experimental covariates that may affect both the intensities and locations of the peaks and spots in the data. I will demonstrate how to identify regions of the functions that are differentially expressed across experimental conditions, in a way that takes both statistical and clinical significance into account and controls the Bayesian false discovery rate to a pre-specified level. I will also discuss how to use this framework as the basis for classifying future samples based on their proteomic profiles in a way that can also combine information across multiple sources of data, including proteomic, genomic, and clinical. Improvements of the modeling framework that result in more robust inference will be discussed, and these methods will be applied to a series of proteomic data sets from cancer-related studies.