

Applications of Wavelet-Based Functional Mixed Models to Proteomics and Genomics Data

Jeffrey S. Morris

Department of Biostatistics, The University of Texas MD Anderson Cancer Center

E-Mail: jefmorris@manderson.org

Abstract: Various genomic and proteomic assays yield high dimensional, irregular functional data. For example, MALDI-MS yields proteomics data consisting of one-dimensional spectra with many peaks, 2D gel electrophoresis and LC-MS yield two-dimensional images with spots that correspond to peptides present in the sample, and array CGH or SNP chip arrays yield one-dimensional functions of copy number information along the genome. In this talk, I will discuss how to identify candidate biomarkers for various types of proteomic and genomic data using 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. Time allowing, I will also demonstrate how to use this framework as the basis for classifying future samples based on their proteomic smf genomic profiles in a way that can also combine information across multiple sources of data, including proteomic, genomic, and clinical. These methods will be applied to a series of proteomic and genomic data sets from cancer-related studies.