ABSTRACT

A central problem in genomic research is the identification of genes and pathways that are involved in diseases or perturbed during a biological process. Many methods have been developed for identifying genes in regression frameworks. The genes identified are often linked to known biological pathways through gene set enrichment analysis in order to identify the pathways involved. However, most of the procedures for identifying the biologically relevant genes do not utilize the known pathway information. In this talk, I present hidden Markov random field (HMRF)-based methods for identifying genes and subnetworks that are activated or perturbed by diseases or biological processes, where the latent gene differential expression states are modeled by a discrete Markov random field. Simulation studies indicated that the methods are effective in identifying genes and subnetworks that are related to disease and have higher sensitivity and lower false discovery rates than the commonly used procedures that do not use the pathway structure information. I will demonstrate these methods by analyzing a breast cancer gene expression dataset to identify the modules related to cancer metastasis and a systemic immune response time course gene expression dataset to identify the subnetworks involved in human immune response to endotoxin.