Incorporating biological knowledge of genes into statistical analysis of genomic data

Wei Pan

Division of Biostatistics. School of Public Health. University of Minnesota. Minneapolis, MN 55455, USA
weip@biostat.umn.edu

In standard statistical methods, genes are treated independently and identically a priori, which ignores biological knowledge on genes in terms of their functions and their inter-relationships. We first review some of our work involving use of gene functional groups, then we discuss further extensions to incorporating gene network information into analysis. The methods cover detecting differential gene expression or DNA-protein interaction, clustering genes based on their expression profiles, and tumor classifications with expression profiles. A common theme of the underlying statistical models is the finite Gaussian mixture model.