Abstract:
This talk addresses the issue of genomic data integration across multiple platforms on the same samples. Making biological insights with such ‘vertically integrated’ genomic data is one of the paramount questions in science and medicine, where such data are quite commonplace. We will discuss the role of hypothesis testing in such a setting. Then we will describe methodology for integrating such data using multivariate extensions of the popular Benjamini-Hochberg (B-H) procedure. In particular, we will exploit two representations of the B-H procedure that correspond to properties of the uniform distribution. This will lead to two classes of multivariate B-H extensions: one motivated by geometric considerations, the other by computational feasibility using kernel machine methodology. The methods are multivariate in nature and are flexible in the nature of the correlation across biological levels. We will illustrate the methodology with simulated and real data examples.