Abstract

With the radical breakthrough in biotechnology, high throughput “omics” data are routinely generated with no limit. These data present unprecedented opportunities in disentangling the genetic secret of complex diseases, while also present daunting challenges in statistical modeling and inference. In the last few years, we have witnessed significant advancement in statistical methodology development for genetic/genomic data analysis, among which kernel based methods have becoming increasingly popular, owing to the advantage in handling nonlinear relationship. In this talk, I will first introduce some background, then focus on two kernel based methods for genomic data analysis: 1) a kernel based testing method to assess gene set association, from a testing perspective; and 2) a kernel fusion method for heterogeneous omics data integration, from a prediction perspective. The utility of the methods will be demonstrated through both simulation and real data analysis.