

## **Multi-scale approaches for analyses of functional phenotypes arising from high-throughput sequencing assays**

Heejung Shim, University of Chicago

Identification of differences between multiple groups in molecular and cellular phenotypes measured by high-throughput sequencing assays is frequently encountered in genomics applications. For example, common problems include testing for association between genetic variants and gene expression using RNA-seq data and testing for differences in chromatin accessibility across tissues/conditions using DNase-seq or ATAC-seq data. These high-throughput sequencing data provide high-resolution measurements on how traits vary along the whole genome in each sample. However, typical analyses fail to exploit the full potential of these high-resolution measurements, instead aggregating the data at coarser resolutions, such as genes, or windows of fixed length. In this talk, I will present two multi-scale approaches that more fully exploit the high-resolution data. In the first part of my talk, I will introduce a wavelet-based approach and demonstrate that the proposed wavelet-based approach has more power than simpler window-based approaches in identification of genetic variants associated with chromatin accessibility. I will also illustrate how the estimated shape of the genotype effect can help in understanding the potential mechanisms underlying the identified associations. The second part will discuss potential limitations of the wavelet based approach in analyses of data sets with small sample sizes or low sequencing depths. To address these issues, I will present another approach that models the count nature of the sequencing data directly using multi-scale models, and demonstrate that the proposed models have substantially more power than the wavelet-based approach in analyses of data sets with small sample sizes or low sequencing depths. While we developed these methods with specific applications to sequencing data in mind, these methods have natural applications for analysis of many functional phenotypes.