

## Sparse learning methods for dissecting the genetic control of biological systems

Since the completion of genome sequencing projects for various organisms including human and other model organisms, the fundamental goal of research in computational genomics, systems biology, and genetics has been to gain a complete understanding of how the instruction sets encoded in genomes get executed within a cell system and organism. The recent advances in the high-throughput technology such as next-generation sequencing technology have allowed the researchers to collect a large amount of data for the genomes and various other aspects of a cell system. Such datasets hold the key to understanding the detailed mechanisms of the genetic control of a biological system and further deepening our knowledge of cell biology with a potential application to medicine. In this talk, I will present statistical machine learning methods that we have developed for learning from high-dimensional genomic data to dissect the genetic control of biological systems. I will focus on sparse learning methods that range from sparse regression methods to sparse probabilistic graphical models, and describe how such methods can be used to effectively extract complex epistatic and pleiotropic interactions among various entities in a biological system. In addition, I will discuss efficient optimization algorithms for learning these statistical models that allow for analysis of large-scale genome-wide datasets. Using yeast genotype and gene-expression dataset, I will demonstrate how our methods can lead to new insights into the activities of genes in a cell as well as the perturbations of gene expressions by genetic variation.