Dr. Jason Ernst, University of California, California

Computational Methods for Large-scale Detection and Dissection of Human Regulatory Elements

Abstract:

Understanding the human genome sequence and in particular the vast non-coding regions is a central challenge for modern molecular biology with profound implications towards understanding the genetic basis of disease. While understanding the genome by directly reading the primary DNA sequence is extremely challenging, the presence of epigenetic marks on top of the sequence holds great promise to aid our understanding of the genome. Technological advances in sequencing has made it possible in a single experiment to generate tens of millions of data points on the location of a epigenetic mark across the genome in a specific cell type, which is raising a number of computational challenges and opportunities. In this talk I will first describe a method that I previously developed, ChromHMM, that learns de novo combinatorial and spatial patterns from maps of multiple epigenetic marks using a multivariate hidden Markov model. These patterns correspond to different classes of genomic elements, which I have then used to provide a cell type specific annotation of the human genome. I will then describe a combined computational modeling and experimental approach that in high-throughput can test putative regulatory elements of interest identified based on epigenomics patterns and identify within them at high resolution bases activating or repressing gene expression. Finally, I will describe a new method, ChromImpute, to impute maps of epigenetic marks that I have applied in the context of the Roadmap Epigenomics project to computationally predict over 4000 epigenomic datasets vastly accelerating the coverage of the human epigenome while providing overall more robust maps than have been obtained experimentally.