DISSERTATION
DEFENSE

Statistical and Machine Learning Methods in the Studies of Epigenetics Regulation.

Tianlei Xu

Emory University

Abstract: Rapid development of next generation sequencing technologies produces a plethora of large-scale epigenome profiling data. Given the quantity of available epigenome datasets, obtaining a clear and comprehensive picture of the underlying regulatory network remains a challenge. The multitude of cell type heterogeneity and temporal changes in the epigenome make it impossible to assay all epigenome events for each type of cell. Computational model shows its advantages in capturing intrinsic correlations among epigenetic features and adaptively predicting epigenome marks in a dynamic scenario. Current progress in machine learning provides opportunities to uncover higher level patterns of epigenome interactions and integrating regulatory signals from different resources. My works aim to utilize public data resources to characterize, predict and understand the epigenome-wide regulatory relationship. The first part of my work is a novel computational model to predict in vivo transcription factor (TF) binding using base-pair resolution methylation data. The model combines cell-type specific methylation patterns and static genomic features, and accurately predicts binding sites of a variety of TFs among diverse cell types. The second part of my work is a computational framework to integrate sequence, gene expression and epigenome data for genome wide TF binding prediction. This extended supervised framework integrates motif features, context-specific gene expression and chromatin accessibility profiles across multiple cell types and scale up the TF prediction task beyond the limits of candidate sites with limited known motifs. The third part of my work is a novel computational strategy for functional annotation of non-coding genomic regions. It takes advantage of the newly emerged, genome-wide and tissue-specific expression quantitative trait loci (eQTL) information to help annotate a set of genomic intervals in terms of transcription regulation. This method builds a bridge connecting genomic intervals with biological pathways and pre-defined biological-meaningful gene sets. Tissue specificity analysis provides additional evidence of the distinct roles of different tissues in the disease mechanisms

Tuesday, April 10, 2018, 10:00 am
Claudia Nance Rollins Bldg. Rm 1036

Steve Qin

MATHEMATICS AND COMPUTER SCIENCE
EMORY UNIVERSITY