Speaker: Christina Leslie

Affiliations: Sloan Kettering Institute

Title: Advances in predictive models for regulatory genomics

Abstract: The last five years have brought notable successes in the application of machine learning approaches, and especially deep learning models, to problems in regulatory genomics. Sequence models based on convolutional neural networks are widely used to interpret DNA motif grammars that explain epigenomic signals like transcription factor occupancy and chromatin accessibility. Other models seek to predict gene expression output from genomic sequence context and/or epigenomic data in order to yield insights into gene regulation. We will discuss new work in our group to overcome limitations in current sequence and gene regulatory models by exploiting 3D genomic architecture, single cell genomics, and natural genetic variation. We will first describe HiC-DC+, our new computational tool for identifying significant and differential 3D interactions from Hi-C and HiChIP data sets, and we will show how to use 3D connectivity in a graph attention network framework to learn predictive models of gene regulation called GraphReg. We will also describe efforts to move predictive models into the single cell realm using scATAC and multiome data sets and to exploit natural genetic variation by training on allele-specific epigenomic data in F1 hybrid mouse models.