Hi-C Landscape of Rhabdomyosarcoma

Meng Wang

Human genome does not exist in a linear fashion. Instead, there are loops and folds, allowing loci far away from each other in genomic distance to be in close physical proximity. It has been shown that the 3D organization of the human genome plays an important part in regulation of gene expression as well as cancer etiology. With Hi-C, it is possible to study the 3D organization of the human genome by constructing a genome-wide contact map.

Not only is Hi-C sequencing library preparation more technically demanding than other high-throughput sequencing technologies, but analysis of Hi-C data also faces unique hurdles. Hi-C data are very large sparse matrices. Classification and pattern recognition of such data are difficult. Furthermore, for cancer research, Hi-C contact matrices violate basic assumptions made during normalization. While many alleviating methods have been proposed to varying degrees of success, the current standards of the Hi-C analysis leave much room for improvement. In this presentation, I will show an application of Hi-C to rhabdomyosarcoma, focusing on the challenges and opportunities in Hi-C data analysis and what can statisticians contribute to mine Hi-C data better.