

GENOME DATA ANALYSIS, PROTEIN FUNCTION
AND STRUCTURE PREDICTION
BY MACHINE LEARNING TECHNIQUES

Renzhi Cao

Dr. Jianlin Cheng, Dissertation Supervisor

ABSTRACT

Genome inside a cell is consist of double stranded DNA sequences, and some special region, so called gene, can be encoded into protein which build the foundation of an organism. Understand the genome structure, protein function and structure is crucial for the quality of our lived. Because of the limitation of current experimental techniques, it is difficult to determine genome structure, protein function and structure. In this research, I analysis the genome conformation data by Hi-C technique, and construct gene-gene interaction networks. The analysis result shows that gene pairs which are spatially interacted tend to share similar function. I apply this finding to predict gene function from its neighbor's gene, and also apply data mining techniques to make protein function prediction. To predict protein structures, I generate different properties of predicted protein structure, and apply machine learning technique for protein structure prediction.