数学与系统科学研究院

计算数学所学术报告

报告人： 刘丙强 教授

（ 山东大学 ）

报告题目：

Systematic Prediction of Regulatory Motifs from Human ChIP-Sequencing Data Based on a Deep Learning Framework

邀请人： 张硕 副研究员

报告时间： 2019 年 11 月 15 日（周五）

下午 **15:00-16:00**

报告地点： 科技综合楼三层

301 报告厅

Abstract：

Identification of transcription factor binding sites (TFBSs) and cis-regulatory motifs (motifs for short) from genomics datasets, provides a powerful view of the rules governing the interactions between TFs and DNA. Existing motif prediction methods however, are limited by high false positive rates in TFBSs identification, contributions from non-sequence-specific binding, and complex indirect binding mechanisms. High throughput next-generation sequencing data provides unprecedented opportunities to overcome these difficulties, as it enables extraction of a full view of the TF's binding activities on the genome level. Meanwhile, it brings new computational and modeling challenges in high-dimensional data mining and heterogeneous data integration. To improve TFBS identification and novel motifs prediction accuracy in the human genome, we developed an advanced computational technique based on deep learning (DL) and high-performance computing, named DESSO. DESSO utilizes deep neural network and binomial distribution to optimize the motif prediction. Our results showed that DESSO outperformed existing tools in predicting distinct motifs from the 690 in vivo ENCODE ChIP-sequencing (ChIP-seq) datasets. We also found that protein-protein interactions (PPIs) are prevalent among human TFs, and a total of sixty-one potential tethering binding were identified among the 100 TFs in the K562 cell line. DESSO was also applied to DNA shape features and found that (i) shape information has a competitive predictive power for TF-DNA binding specificity; and (ii) identified shape motifs are substantially recognized by human TFs and contribute to the interpretation of TF-DNA binding in the absence of sequence recognition. The developed tool and subsequent analyses will improve our understanding of how gene expression is controlled by the underlying regulatory systems.

欢迎大家参加！