中国科学院数学与系统科学研究院
Academy of Mathematics and Systems Science, CAS

生物信息学研究中心

Center of Bioinformatics

学术报告

题　目： Discovering biological progression underlying high dimensional data

报告人： Peng Qiu

Department of Bioinformatics and Computational Biology, Univ of Texas M.D. Anderson Cancer Center

时　间： 7 月 11 日（星期三）下午 4:00

地　点： 思源楼 1013 室

摘要：

We present a novel computational approach, Sample Progression Discovery (SPD), to discover patterns of biological progression underlying high-dimensional datasets. In contrast to the majority of microarray data analysis methods which focus on identifying differences between sample groups (i.e. normal vs. cancer, treated vs. control), SPD aims to identify an underlying progression among individual samples, both within and across sample groups. This is essentially a new way of asking questions. The traditional analyses ask the following question: what is the different between A and B. In this talk, I am going to ask a different question: how did A become B, or how did one biological sample/phenotype go through gradual changes and eventually progress into another phenotype. In cancer studies, this is to ask: how did normal samples go through progressive changes and eventually become cancerous. The SPD method is designed to address this progression question. To demonstrate the utility of SPD, we applied it to gene expression datasets of cell cycle time series, B-cell differentiation, mouse embryonic stem cell differentiation, and prostate cancer. Each of these datasets is associated with a known biological progression. The known progression was hidden from the algorithm and was only used to validate the results. When applied to these datasets, SPD successfully recovered the underlying progression and genes that are associated with the progression. We will also discuss cases where SPD fails. For example, when applied to dataset without any underlying progression, SPD degenerates into a clustering tool.