Visualization and data mining of tremendous cancer transcriptome data

Zefang Tang  
Peking University, China

The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTeX) projects produced RNA-Seq data for tens of thousands of cancer and non-cancer samples, providing an unprecedented opportunity for data mining, cancer drug target discovery and data visualization. In recent years, promising cancer drugs including panitumumab and bevacizumab have been developed that inhibit cancer cells by selectively targeting over-expressed EGFR or VEGF genes in cancer cells, while leaving normal cells unharmed. Genetic alterations will influence gene expression directly or indirectly. It is a frequently used strategy to discover candidate cancer drug targets through the finding of cancer specific expressed genes. This study aims to investigate normalization methods for integrating different expression datasets, explore effective approaches to obtain differentially expressed genes, profile the prognostic genes and transcripts in survival analyses, characterize the distribution of cancer specific genes or transcripts, and analyze their biological functions. Meanwhile, we will develop tools for visualizing integrated expression data, with the aim to disseminate such data to the wide research community. We also plan to find useful biomarkers for early diagnosis. Finally, by investigating the association between genetic alterations and over-expression, we aim to elucidate the underlying genetic mechanisms of differentially expressed genes.

tangzefang@pku.edu.cn