Model-based clustering of time-course RNA-seq data

Man-Kee M. Chu*
The University of Western Ontario, London, Ontario, Canada mchu5@stats.uwo.ca

Wenqing He
The University of Western Ontario, London, Ontario, Canada whe@stats.uwo.ca

The next generation sequencing technology (RNA-seq) provides absolute quantification of gene expression using counts of read. Transcriptome studies are switching to rely on RNA-seq rather than microarrays since RNA-seq has higher sensitivity and dynamic range, with lower technical variation and thus higher precision than microarrays. Limited work has been done on expression analysis of longitudinal RNA-seq data to account for the time-dependence nature of the count data with over-dispersion property. Functional clustering is an important method for examining gene expression patterns and thus discovering co-expressed genes to better understand the biological systems. We propose a model-based clustering method for identifying gene expression patterns using time-course RNA-seq data. A time-course genomic dataset is employed for illustration.

Key Words: Expectation-Maximization algorithm, longitudinal experiments, over-dispersed count data, mixture model