UCLA

BIOMATH

Using model-based methods to quantify exon-level gene expression from RNA-seq data

Zhaohui Steve Qin, Ph.D.

Associate Professor
Department of Biostatistics and Bioinformatics
Rollins School of Public Health
Emory University

Thursday, Sept. 29, 2011
4:00 PM
23-105 Center for the Health Sciences (CHS)

ABSTRACT:
RNA sequencing (RNA-seq) is a powerful new technology for mapping and quantifying transcriptome using ultra high-throughput next generation sequencing technologies. Using deep sequencing, gene expression levels of all transcripts including novel ones can be quantified digitally. Although extremely promising, the massive amounts of data generated by RNA-seq, substantial biases, and uncertainty in short read alignment pose daunting challenges for data analysis. In particular, large base-specific variations and between-base correlations make simple approaches, such as those that use averaging to normalize RNA-seq data and quantify gene expressions, ineffective. In this study, we propose a model-based method to characterize base-level read coverage within each exon. The underlying expression level is included as a key parameter in this model. Since our method is capable of capturing local genomic features that affect read coverage profile throughout the exon, we are able to obtain improved quantification of the true underlying expression levels.

Host: Marc Suchard, Ph.D.

To receive e-mail seminar notices, contact David Tomita (dtomita@biomath.ucla.edu)