Upgraded Genomic Association Analysis via New Software Platform

October 21, 2013

ND Weber

The field of genomics is rapidly evolving, and inherent in this evolution is the development of better and more advanced computational capabilities. A research group led by Dr. Raphael Gottardo in the Vaccine and Infectious Disease Division is working to improve the software used in an area of gene regulation studies called eQTL mapping. Their work will help to upgrade current capabilities for utilizing the ever-expanding amount of sequencing and microarray data to draw conclusions about biological questions.

Genome-wide association studies (GWAS) involve analyzing how variations in DNA affect traits. One aspect of these studies focuses on identifying specific genomic regions that influence gene expression levels. These regions are called expression quantitative trait loci (eQTL). The identification and mapping of eQTL to genes can provide insight into gene regulation and the factors that affect gene expression networks, including gene expression patterns that are associated with disease. Current eQTL mapping software only considers genes individually. However, because of the existence of gene networks comprising large numbers of genes and eQTL 'hotspots' (genomic loci that affect expression of a large number of genes across the entire genome), being able to analyze multiple genes together when analyzing potential eQTLs would provide a more powerful tool for performing these studies. A recent paper published in the journal *Bioinformatics* by the Gottardo laboratory presents a new software package to do just that.

"Advances in assay technology often outpace the development of models to analyze their results," explains lead author Greg Imholte, a graduate student in the Gottardo Lab. "As gene and DNA sequence data sets grow staggeringly large, modeling efforts must simultaneously cope with heavy computational burdens and a compulsion to detect signals in data without suggesting too many false results." The software package they developed to address these issues, referred to as iBMQ, "balances these two opposing desires by leveraging modern computational techniques and the flexible Bayesian modeling paradigm." The Bayesian method of statistical modeling allows for information to be shared across markers and/or genes in order to increase the power to detect eQTLs.

A main advantage that iBMQ offers is that all the gene expression data and data on DNA sequence variations, or SNPs (single nucleotide polymorphisms), can be analyzed concurrently by incorporating them into the same model. The net output reveals an "increase in the power to detect eQTL hotspots that play an important role in the dynamic and global nature of transcriptional regulation," explains Imholte.

Naturally, as with all advances in data analysis software, there is still room for growth. Mr. Imholte envisions an area of improvement coming from "the incorporation of prior knowledge on the functional nature of specific genomic loci (e.g. transcription factor, nucleosome occupancy) that could be used to improve the detection of functional eQTLs."

Imholte GC, Scott-Boyer MP, Labbe A, Deschepper CF, Gottardo R. iBMQ: a R/Bioconductor package for integrated bayesian modeling of eQTL data. *Bioinformatics.* Epub ahead of print, doi: 10.1093/bioinformatics/btt485.
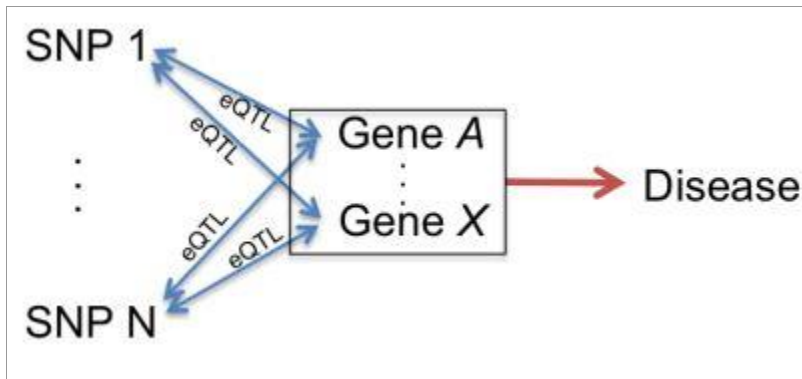

Image adapted from Hao Li Laboratory at The University of California, San Francisco website (http://sherlock.ucsf.edu/)

Expression quantitative trait loci (eQTL) mapping involves analyzing the associations between genetic variations (SNPs) and gene expression. The new software iBMQ offers an improved method for carrying out these analyses.