New tools yield superior genome analysis results

December 1 2015

Scientists from the Icahn School of Medicine at Mount Sinai have developed and publicly released new data analysis software that could help genomics researchers identify genetic drivers of disease with greater efficiency and accuracy. These tools were published yesterday in *PLoS Computational Biology* and on November 25th in *Scientific Reports*, a Nature publication.

MEGENA (for Multiscale Embedded Gene Co-expression Network Analysis) projects gene expression data onto a three dimensional sphere, allowing scientists to study hierarchical organization patterns in complex networks that are characteristic of diseases such as cancer, obesity, and Alzheimer's. Tested on data from The Cancer Genome Atlas (TCGA), MEGENA identified novel regulatory targets in breast and lung cancers, outperforming other co-expression analysis methods.

The second tool, SuperExactTest, establishes the very first theoretical framework for assessing the statistical significance of multi-set intersections and enables users to compare very large sets of data, such as gene sets produced from genome-wide association studies (GWAS) and differential expression analysis. Scientists ran SuperExactTest on existing TCGA and GWAS data, identifying a core set of cancer genes and detecting related patterns among complex diseases. Both tools come from the Multiscale Network Modeling Laboratory led by Bin Zhang, PhD, Associate Professor in the Department of Genetics and Genomic Sciences.

"These tools fill important and unmet needs in genomics," said Dr. Zhang, the senior author of the two papers. "MEGENA will help scientists flesh out novel pathways and key targets in complex diseases, while SuperExactTest will provide a clearer understanding of the genome by comparing a large number of gene signatures."

"Our team is dedicated to crafting high-performance analysis tools and to sharing those resources with the broader genomics community to help us all generate the best possible results," said Eric Schadt, PhD, the Jean C. and James W. Crystal Professor of Genomics at the Icahn School of Medicine at Mount Sinai, and Founding Director of the Icahn Institute for Genomics and Multiscale Biology. "These new tools demonstrate thoughtful and creative solutions to computational challenges faced by scientists around the world, and I look forward to seeing what the community will accomplish with them."

MEGENA and SuperExactTest are available as R packages at Dr. Zhang's website and CRAN (the Comprehensive R Archive Network), a repository of open-source software.

 More information: Won-Min Song et al. Multiscale Embedded Gene Co-expression Network Analysis, *PLOS Computational Biology* (2015). DOI: 10.1371/journal.pcbi.1004574

Minghui Wang et al. Efficient Test and Visualization of Multi-Set Intersections, *Scientific Reports* (2015). DOI: 10.1038/srep16923

Provided by The Mount Sinai Hospital