New software for the visual analysis of genome-wide expression data

Biologists use modern high-density micro arrays to measure the genome-wide expression of exons, the essential parts of alternatively spliced genes. In particular, exons of many genes encode protein variants involved with important cellular processes. While the measured differences between the variants are often quite small, their distinct biological effects can be profound. Protein variants from the same gene might even have opposing functional consequences. Therefore, scientists from the Max Planck Institute for Informatics, Saarbrücken, Germany, and the Gladstone Institute of Cardiovascular Disease, San Francisco, CA, USA, developed innovative software to analyze exon expression data.

The new software package for the analysis of the massive data generated by exon micro arrays consists of the programs AltAnalyze (www.altanalyze.org) and DomainGraph (www.domaingraph.de) and is amenable even to biologists without programming knowledge and experience in bioinformatics. It is the first software that allows researchers not only to perform statistical data analysis, but also to evaluate the biological implications of alternative splicing on cellular processes.

“One once we realized the need for the joint statistical and visual analysis of this type of expression data, we decided to combine our efforts and work towards a free software product that would be both user-friendly and comprehensive in its scope,” says Mario Albrecht, research group leader at the Max Planck Institute and this year's recipient of the international HUPO Early Career Investigator Award for outstanding young proteome researchers.

The conceptual aim of the software is to provide researchers with an easily accessible way of producing and interpreting statistical results for their large expression data files. Thus, the analysis results can be visualized in the context of molecular networks and pathways, supporting different levels of molecular details so that the biological user is not overwhelmed by the data deluge. “Rather than just providing statistical results as long and cryptic tables, our software empowers users to visually explore the functional consequences of the identified protein variants in mammalian cells. Researchers can readily study the resulting biological differences such as altered protein function and pathway interactions together with the measured exon expression values,” explains Dorothea Emig, research scientist at the Max Planck Institute.

“While mainstream biology has embraced the analysis of gene expression, the extent of alternative splicing has not been researched yet at genomic scale. However, this basic mechanism can have tremendous influence on how cells respond under normal conditions and disease. Our software tools help bringing this kind of science to the forefront of biology research,” adds her cooperation partner Nathan Salomonis, postdoctoral fellow at the Gladstone Institute. Recently, this software provided Salomonis and colleagues with new predictions on protein variant functioning and targeting by small inhibitory RNAs, which were further validated in embryonic stem cells. “With the software results in hand, the user is presented with functional hypotheses that can be readily tested in further lab experiments. They will have a much greater likelihood of being relevant to the molecular processes of interest”, notes Bruce Conklin, senior investigator at the Gladstone Institute and professor of medical genetics at the University of California, San Francisco.

Original publications:

Contact:
Dr. Mario Albrecht
Max Planck Institute for Computer Science
Campus E1.4
66123 Saarbrücken
Germany
Tel.: +49-681-9325-327
Fax: +49-681-9325-399
E-Mail: mario.albrecht@mpi-inf.mpg.de
WWW: http://medbioinf.mpi-inf.mpg.de

Samir Hammann
Max Planck Institute for Computer Science
Campus E1.4
66123 Saarbrücken
Germany
Tel: +49-681-9325-454
E-Mail: s.hammann@mpi-inf.mpg.de