A Tool for Quantitative Analysis of Proteomics Data

Ashoka D. Polpitiya[1*], Navdeep Jaitly[2], Konstantinos Petritis[1]

1. Center for Proteomics, Translational Genomics Research Institute, Phoenix, AZ 85004.
2. Department of Computer Science, University of Toronto, Toronto, ON M5S 3G4.
*Contact author: ashoka@tgen.org

Keywords: proteomics, normalization, statconnDCOM, R,C#

Proteomics involves the study of protein content of an organism, a complement of its genome. Mass spectrometry coupled with liquid chromatography techniques are the most popular methods used for proteomics studies. Although there are number of tools available for high-throughput microarray data processing (Gentleman et al. 2004), they are not ideally suited to address the issues specific to proteomics data. For example, in the most common approach called the "Bottom-up" proteomics where proteins are first enzymetically digested to obtain smaller peptides that are easier to detect in a mass spectrometer, inferring the protein quantities from the observed peptides is a unique challenge. Another major issue is the extent of missing values that is largely due to the amount of species near the threshold for detection leading to unbalanced datasets.

Inferno is a free, open source statistical tool designed to address the unique challenges associated with large scale proteomics studies. The graphical user interface (GUI) of Inferno is implemented in C# language and the core algorithms are implemented in R (R Development Core Team, 2009). statconnDCOM (Baier et al., 2010) is used for the connectivity between the .NET environment and R. Inferno runs on Microsoft Windows platform.

Inferno features many statistical plots such as boxplots, histograms, QQ plots, and correlation diagrams. A set of normalization algorithms such as LOWESS and linear regression, is also implemented for removing any systematic variation in the data. It also presents few methods to infer protein abundances from the observed peptide abundances and a comprehensive ANOVA scheme based on the **car** package (Fox, 2002), for selecting significant proteins in an experiment.

Inferno is available at `http://inferno4proteomics.googlecode.com`.

References

R Development Core Team (2009). R: A Language and Environment for Statistical Computing, `http://www.R-project.org`.

Baier T., Neuwirth E. (2010). Statconn DCOM, `http://rcom.univie.ac.at/`.

Gentleman RC, et al. (2004). Bioconductor: open software development for computational biology and bioinformatics, Genome Biol., 5(10):R80.

Fox, J. (2002). An R and S-PLUS Companion to Applied Regression. Sage Publications, Thousand Oaks, CA, USA.