Overview

- Quantitative proteomics has become increasingly effective in understanding the biology and biomarkers for diseases.
- Challenges in quantitative proteomics:
  - Systematic variations among technical and biological replicate measurements
  - Inference of protein abundances from the observed peptide abundances
  - Undetected peptides leading to "missing values"
- DAnTE (Data Analysis Tool Extension) is designed to address these issues featuring:
  - Normalization methods
  - Missing value imputation algorithms
  - Peptide to protein rollup methods
  - Statistical plots
  - Hypothesis testing scheme that can handle unbalanced data and random effects

Removing Systematic Variation – “Normalization”

Factors

Factors capture the experimental design via fixed and random effects. This information is later used in normalization, imputation, and hypothesis testing methods in DAnTE.

Data Normalization Methods

- Robust linear regression
- Lowess
- Quantile method
- Global intensity adjustment using Median Absolute Deviation
- Central tendency adjustment

Normalizing data to remove systematic variation

DAnTE Features

Multiple Analysis
  - Biological conditions
  - Biological replicates
  - Technical replicates

Factors

Factors with an FDR of 5% include 361 proteins (ANOVA with an FDR of 5%)

Significant Proteins

Figure 1. Analysis flow in DAnTE

Figure 2. Analysis flow in DAnTE

Figure 3. Normalizing data to remove systematic variation

Protein Quantitation

Protein with 9 detected peptides:

hmg ampl WT

Raw peptide abundances vs. dataset (for 1 protein)

Scalped peptide abundances for this protein’s 28 peptides

References


Acknowledgements

Significant portions of the work were performed in the Environmental Molecular Science Laboratory, a DOE/BER national scientific user facility at PNNL in Richland, Washington. PNNL is operated for the DOE by Battelle under contract DE-AC05-76RLO-1830.

Portions of this research were supported by the National Institute of General Medical Sciences (NIGMS, Large Scale Collaborative Research Grants U54 GM-62119-02), the National Center for Research Resources (RR18522), the National Institute of Allergy and Infectious Diseases NIH/DHHS (through interagency agreement Y1-AI-4894-01). Samples were analyzed using capabilities developed under the support of the National Center for Research Resources and the U.S. Department of Energy Biological and Environmental Research (DOE/BER).

DAnTE is designed as a complete downstream solution for quantitative proteomics, addressing key issues, such as the incomplete nature of the data and inference of protein abundances from the observed peptides. We are continuously investigating new algorithms that improve the quality of the analysis and the tool is readily extendable.

DAnTE is available at http://omics.pnl.gov/software

CONTACT: Ashoka D. Polpitiya, D.Sc.
Biological Sciences Division, K8-98
Pacific Northwest National Laboratory
P.O. Box 999, Richland, WA 99352
E-mail: ashoka.polpitiya@pnl.gov