

# Quantitative Analysis Methods in Proteomics

Ashoka D. Polpitiya, Wei-Jun Qian, Vladislav A. Petyuk, Joshua N. Adkins, Gordon A. Anderson, Richard D. Smith  
Pacific Northwest National Laboratory, Richland, WA



Pacific Northwest  
NATIONAL LABORATORY

## 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<sup>1</sup> (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

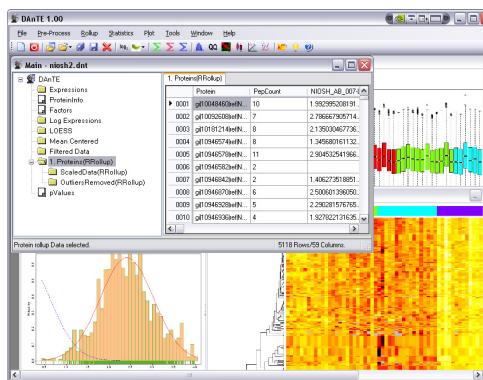


Figure 1. DAnTE: Data Analysis Tool Extension Software

## DAnTE Features

### Multiple Analysis

- Biological conditions
- Biological replicates
- Technical replicates

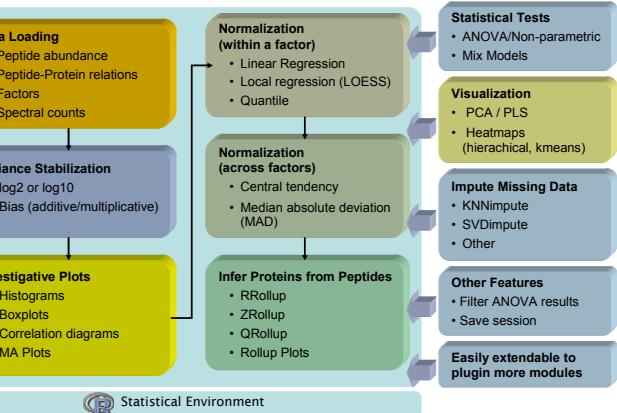
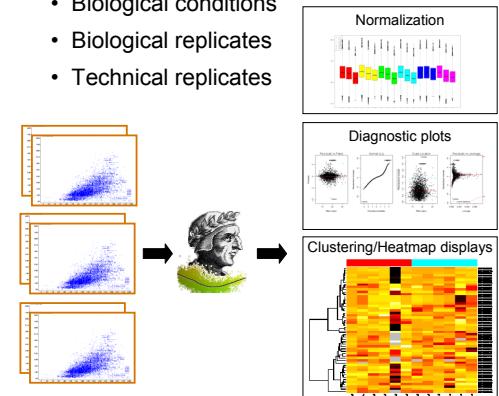


Figure 2. Analysis flow in DAnTE

## 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.

## Removing Systematic Variation – “Normalization”

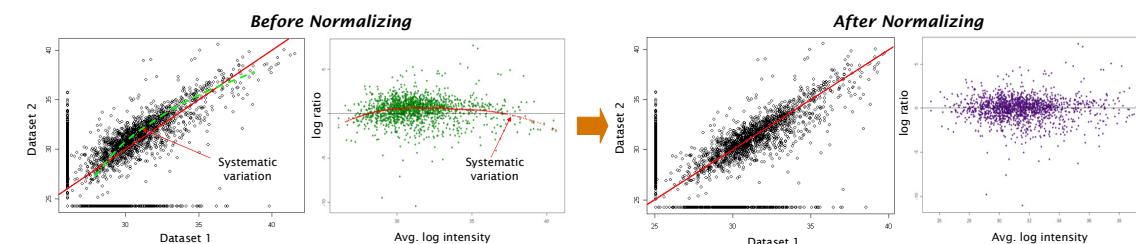


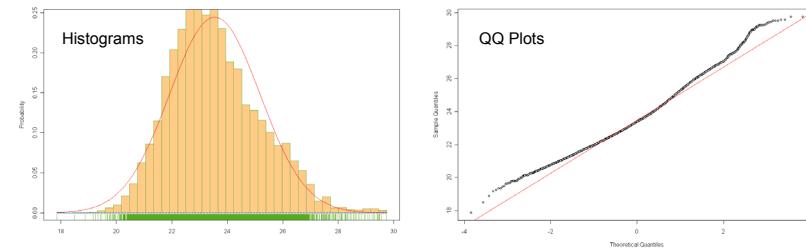
Figure 3. Normalizing data to remove systematic variation

## Data Normalization Methods

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

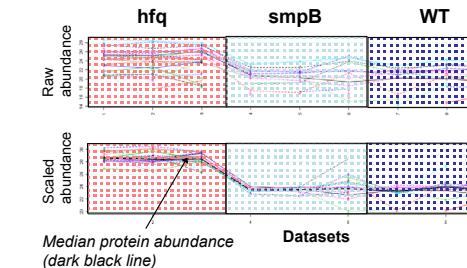
## Statistical Plots

- Histograms
- QQ plots
- Box plots
- Correlation diagrams
- MA plots



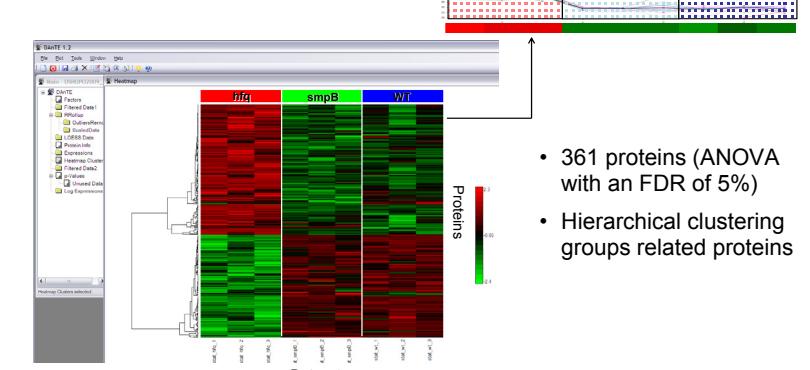
## Protein Quantitation

Protein with 9 detected peptides:



Raw peptide abundances vs. dataset (for 1 protein)  
Scaled peptide abundances for this protein's 28 peptides

## Significant Proteins



- 361 proteins (ANOVA with an FDR of 5%)
- Hierarchical clustering groups related proteins

## Summary

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>

Biological MS Data and Software Distribution Center

**Site Links**

Home  
Browse Available Data  
Search Available Data  
Download Software  
Register New Account  
Research Sponsors  
Posters and Presentations

**DAnTE**  
DAnTE helps users to perform various downstream analysis, data reduction, and data comparison tasks. DAnTE includes a graphical environment for visualizing the data during the processing steps. DAnTE can associate one or more “factors” with each dataset, which are then taken into consideration during the data processing. The statistical methods are implemented in R, and the graphical user interface is written in C#, using the RDLCOM server for connectivity.  
Features available in DAnTE include:

- Comprehensive data loading mechanism
  - Tabular data from CSV, Excel files (either abundances or expression ratios)
  - SQL database tables for spectral count analysis
- Statistical methods for spectral count analysis
  - Diagnostic plots
    - Histograms, boxplots, correlation plots
    - Peptide-to-protein ratio plots

## Acknowledgements

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 NIH National Center for Research Resources (RR18522), Laboratory Directed Research and Development (LDRD) program (V-JC) at Pacific Northwest National Laboratory (PNNL), and 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 NIH National Center for Research Resources and the U.S. Department of Energy Biological and Environmental Research (DOE/BER). 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.

## References

- AD Polpitiya, et al. *Bioinformatics* 2008, 24: 1556-1558.
- SJ Callister, et al. *J Proteome Res* 2006, 5: 277-286.

**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