

DAnTE: a statistical tool for quantitative analysis of proteomics data

Ashoka D. Polpitiya, Wei-Jun Qian, Navdeep Jaitly, Vladislav A. Petyuk, Joshua N. Adkins,
David G. Camp II, Gordon A. Anderson and Richard D. Smith
Pacific Northwest National Laboratory

Overview

- DAnTE (Data Analysis Tool Extension) is a new statistical tool designed to address challenges associated with quantitative bottom-up, shotgun proteomics data. This tool has also been demonstrated for microarray data and can easily be extended to other high-throughput data types.
- DAnTE features selected normalization methods, missing value imputation algorithms, peptide to protein rollup methods, an extensive array of plotting functions, and a comprehensive hypothesis testing scheme that can handle unbalanced data and random effects. The Graphical User Interface (GUI) is designed to be very intuitive and user friendly.
- DAnTE is available for download at <http://ncrr.pnl.gov/software/>

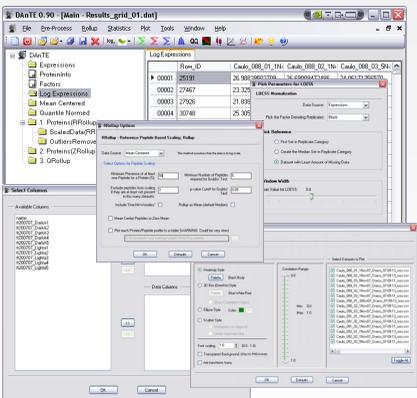


Figure 1. Various screenshots of DAnTE



DAnTE Features

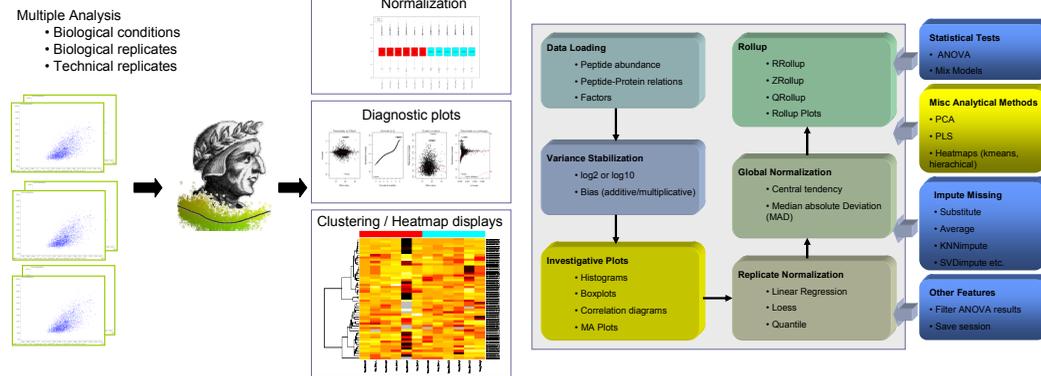


Figure 2. Analysis flow in DAnTE

Data loading:

DAnTE employs a sophisticated method to load data from any file that stores tabular data (csv, tab-delimited or Excel format). DAnTE extracts all peptide, protein, and dataset annotations from the input file once the relevant columns are selected using the data loading dialog box.

Investigative Plots:

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

Factor Definitions:

Factors are used to capture the fixed and random effects in experimental design. This information is vital in normalization, imputation, and hypothesis testing methods in DAnTE. Factors can either be declared once the data is loaded or be loaded from a flat file.

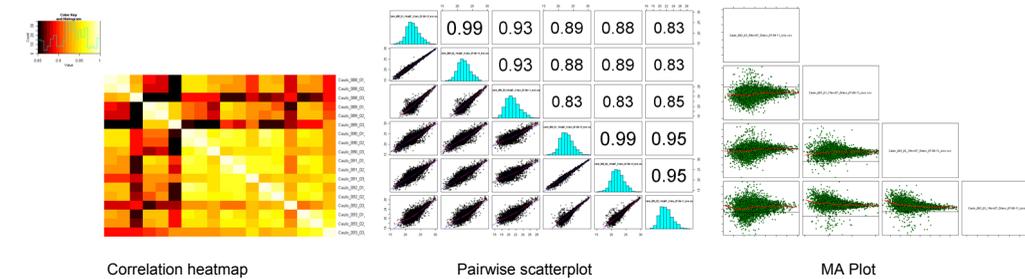


Figure 3. Diagnostic plots in DAnTE

Data normalization methods:

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

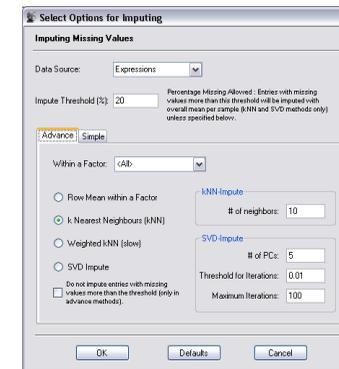


Figure 5. Options for imputing missing values

Peptide to protein rollup:

- RRollup
- ZRollup
- QRollup

Hypothesis testing

- ANOVA
 - Balanced/unbalanced
 - Fixed/random effects
- Shapiro-Wilk Test for normality
- Non-parametric tests
 - Wilcoxon
 - Kruskal-Wallis

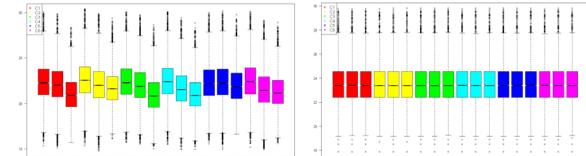


Figure 4. Boxplots before and after normalization: Boxes are colored based on a factor (biological condition). Left plot shows raw data having a bias and the normalization steps can remove it (right).

Missing value imputation:

- Row mean based on a factor
- k Nearest Neighbors
- Weighted kNN
- SVD Impute
- Mean/median of a sample
- Substitute with a constant

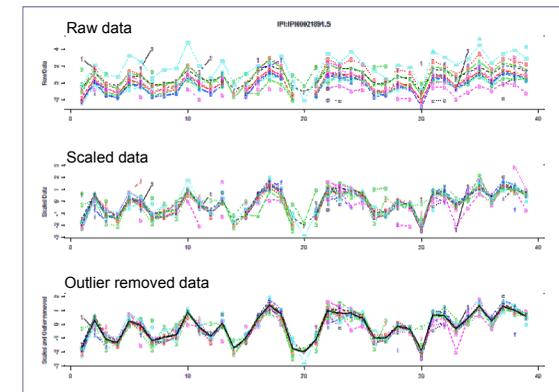


Figure 6. Results from RRollup method

Summary

- DAnTE is a first generation downstream analysis tool that allows collaborators to explore large volumes of data and extract biological knowledge without the aid of bioinformatics staff.
- DAnTE can help users to:
 - remove any systematic bias (normalization)
 - rollup peptides to proteins
 - explore patterns in the data
 - discover features (e.g., biomarkers)