RGtk2Extras and DanteR: Rapid GUI development for an "omics" R package

Tom Taverner1, Ashoka Polpitiya2, Gordon A Anderson1, Richard D Smith1
1 Biological Sciences Division, K-8, Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352
2 Translational Genomics Research Institute, Phoenix, AZ 85004

Abstract
Proteomics research results in large, complex datasets, presenting the researcher with many detected features over hundreds of datasets. The features detected provide evidence of the presence of constituents in the samples, including peptides, proteins, and unidentified features. It can be challenging to normalize and visualize the results of a comprehensive, quantitative "omics" study due to the volume of the data and the need to interactively explore the results to detect changes and patterns. We have developed the DanteR package and some other tools to address these needs.

The DanteR Package
DanteR is an R package with a focus on normalization and statistical analysis of quantitative proteomics studies. Its aim is to provide scientists with an interactive, graphical interface that guides users through the preprocessing/analysis process. It is built entirely in the widely used R language and is open-source and platform independent, which are important considerations for a target audience of scientists using a variety of operating systems.

The RGtk2Extras Package
This package allows graphical front-ends for functions and interactive components to be easily specified by users with minimal R experience but no background in GUI development. The package also includes Rgit2XEd0, an RGtk2-based spreadsheet-like data viewer and editor featuring improved features and capabilities over currently available editor widgets.

DanteR Features and Functionality
- History tab, Object Browser, Table Display
- File I/O: Import Tables, Workflow, Open/Save Session, Export Tables
- Data: Sort or filter by data field or linked metadata field, merge data
- Pre-processing: Linear Regression, LOESS, Quantile Regression, Central Tendency, Imputation
- Extensive help and walkthroughs

• Statistics: ANOVA and GLM, Fold Changes, Fisher Test, Nonparametric Tests, G-Test, Post-hoc Adjustment, Filter Statistically Significant
• Plotting: Scatter Plot, 3D Plot, Boxplots, Histogram, QQ Plot, Venn Diagram, many more
• Explore: Clustering, Pattern Search, PCA/PLS, Dynamic Row Plot – Dynamic Row Plot allows on-the-fly SICL-like queries on linked data table, for example low p-values
• Dynamically updatable Addons menu using SimpleDialogMaker
• Created in close collaboration with biologists and statisticians

Conclusions and Future Work
We have created R functions for GUI and dialog development which we have found to be useful in the rapid development of a proteomics-related R package and front end, DanteR. This includes a toolkit for simple, rapid dialog development and a data table editor, which are included in the RGtk2Extras package.

Given the power of the R language and its broad use in the statistical community, R provides an excellent foundation to build the data integration framework. For example, one extension to DanteR we are planning is integration with components of the Bioconductor project. Integration with DanteR will allow pathway analysis for proteomics data and make use of existing work in genomics for integrative projects.

References

Contact: Thomas.Taverner@pnl.gov