

What's New in BRB-ArrayTools Version 4.3.0 (details given in users guide)

=====

We have made a few major changes in this version. First, we have added the option of importing RNA-Seq data pre-processed with the Galaxy web tools (<https://main.g2.bx.psu.edu/>). Using Galaxy, large data files can be processed, mapped and converted to FPKM estimates and then imported into BRB-ArrayTools for further utilization of a large variety statistical analysis tools on expression data.

Secondly, we have added an importer for the Illumina Methylation data. Methylation data output from the Illumina GenomeStudio software can be imported into BRB-ArrayTools through the Data Import Wizard for further statistical analysis. We have also implemented two tools for finding frequently methylated probes and for correlating methylation and expression data.

In this version, the RServe package is used for communication between Excel-Visual Basic and R code. This removes the dependency of ArrayTools on statconnDCOM and allows 64-bit R to run under all circumstances in a 64-bit operating system.

Visualization tools

A new tool called "Heatmap of data" is provided to generate a heatmap on clustered data to provide users an overview on their data. In addition, a zoomable heatmap in SVG format is generated after running either the "Clustering Genes and Samples" or "Heatmap of data" tool.

Analysis Tools

Added a plug-in to find frequently methylated probes.

Added a plug-in to correlate methylation with expression.

Class Prediction: Added ROC curves for the Compound Covariate Predictor and the Diagonal Linear Discriminant Analysis classifiers.

Lassoed Principal Components: the HTML output now includes an expression table.

Random forest plug-in: If the user's computer has a multi-core processor, parallel computing will be used.

PAM: Enhanced by including the shrunken centroids in the HTML output file and allowing the user to select a random seed for the permutation test.

Lasso logistic regression: Added to output the predicted probabilities for test samples. The user can specify the number of genes to be retained in the model.

Survival risk prediction: Computes ROC curves (sensitivity vs 1-specificity) at landmark time. Added option of specifying the number of genes to be included in the model. Added an option of evaluating statistical significance based on using area under the ROC curve as the test statistic for the permutation test.

csSAM Analysis: Modified code to allow only a subset of samples used in the cell frequency file for the analysis.

Data Import

Implemented a new option to import Illumina methylation data.

Implemented a new option to import RNA-Seq data pre-processed using the Galaxy web tools (<https://main.g2.bx.psu.edu/>)

Modified Visual Basic code to save projects in .xlsx format and remove the limitation of 65k genes in Excel 2007/2010.

Annotations: Added a Utility to allow importing annotation information from an annotated project with the identical chip type.

Data Filtering

Added MicroRNA, protein domain, transcription factor, BROAD C2 genesets to filtering options.