

What's New in BRB-ArrayTools v4.1 and CGHTools v1.2

Visualization tools

New 2-D and rotating 3-D interactive scatterplot tools have been implemented with a variety of features like multi panels, linking plots, highlighting genes based on pathways etc. To view the enhanced graphics, here is a link to the online demo <http://linus.nci.nih.gov/PowerPointSlides/Scatterplot.wmv>

Heatmaps

The clustering heatmaps have been re-designed to handle more genes and arrays. The images have been enhanced with rectangular pixels and class labels have been added. The color palette for the analysis related heatmaps can now be modified.

Analysis Tools

Gene Set Expression Analysis: An optional interaction analysis has been added to find gene sets for which the inter-class differential expression varies among pre-defined groups of samples.

Another new feature is the inclusion of gene sets based on lymphoid signatures from the Staudt lab(<http://lymphochip.nih.gov/signaturedb/>). We have also updated all the existing gene sets within ArrayTools.

Class comparison: The pair-wise option now permits more than two class levels.

Lassoed Principal Components plug-in: We have implemented Witten and Tibshirani's new method for identifying genes whose expression varies among classes, is correlated with a quantitative trait or is correlated with survival time.

Adaboost plug-in: A tool for class prediction using the Adaboost method developed by Freund and Schapire (1996) has been implemented as a plug-in. Classification is based on weighted voting of a set of classification trees.

Data Import

Affymetrix Gene ST Array Importer: A platform specific data importer is provided for human, mouse and rat Gene ST 1.0 arrays.

GenePix importer: The data import wizard can now handle single channel GenePix data.

Custom Annotations: This release permits import of user supplied gene annotations for custom species/arrays.

Annotations: SOURCE annotations can now be imported for 8 different organisms.

Data Filtering

An option is provided for selecting a single probe/probe set for each gene represented on the array.

Utilities

A new utility is provided that obtains drug bank information for all genes in a gene list produced by any BRB-ArrayTools analysis. This provides drugs whose targets include protein products of genes on the specified list.

Genelists are now created for both positive and negative correlations to a specific gene.

The user can now control the heatmap plot options from the preferences option under utilities.

CGHTools

The HaarSeg algorithm is provide as an alternative and faster segmentation method. All segmentation is now performed by loading one sample at a time to improve memory handling for large data sets.

Pathway enrichment analysis can now be performed for mouse as well as human arrays. Support for rat and mouse arrays in GISTIC analysis and in integrated analysis between copy number and expression is now provided.

The identification of frequent copy number aberrations can now run on either arrays of a specified class or on all the arrays.

The general importer can now import individual red and green intensities and compute the corresponding log₂ratios.

The following changes were made since the last v4.1 beta_3 release:

Analysis tools:

- 1) Fixed an error in the PAM prediction analysis tool caused by using -999999 values instead of "NA" for missing data.
- 2) The Volcano plot in class comparison was modified to use $1e-7$ as a threshold for genes with p-values $< 1e-7$. The parallel coordinate plot was modified to fix an error if the random variance model was selected.
- 3) The Adaboost plug-in has been changed to fix a problem of plotting trivial trees (root only) on HTML output. Also fixed a bug that occurred when the random variance model was selected with Adaboost.
- 4) For the Gene set Comparison tool with User-defined gene lists, turned off the option to use a minimum/maximum number of genes.
- 5) Enhanced the output for survival analysis by adding gene symbols to the Kaplan-Meir curves.

Clustering:

- 6) Fixed errors that occurred when reading the results from a saved project for the Zoom and Recolor, List Genes, and Cut Tree functions in Clustering Genes and Samples. Also, added the clustering order of samples to the output.

Importing and Filtering:

- 7) Modified the code to give an appropriate warning message that the spot flag filter will only be applied to non-replicate spots if the "average replicate spots" option was selected.
- 8) Modified the code to recognize the Affymetrix probe-set data file with "- Signal" in the column header.
- 9) Modified the code to handle converting txt files from UNIX to DOS format when the raw data files were placed under a folder with spaces in path/folder name.
- 10) Fixed a bug where some Agilent single channel data was incorrectly recognized as dual channel data during importing.
- 11) Modified the code to turn off normalization when the gene list file for housekeeping gene normalization resulted in no matches with the expression data.

Utilities:

- 12) Updated the Drug Bank utility links to correspond to the latest version DrugBank3.
- 13) Fixed problems related to running the unzip tool for gene lists files.
- 14) Modified the code to handle "---" in the annotation results
- 15) Added the annotation for Rhesus Macaque through SOURCE.

Installation:

16) In this release, users will be required to enter their password/emails after installation.

CGHTools:

- 1) Fixed an indexing error in pathway analysis when MAD (median absolute deviation) factor is selected for gain/loss determination.
- 2) Fixed a bug in importing Illumina CGH data.
- 3) Fixed a bug when "Segment.xls" needs to be written out at the GISTIC step for a project collated prior to v1.2.
- 4) Modified the code to handle an error caused when using a non-English version of Office.