This is a text file named "Readme.doc" in the BRB-ArrayTools installation directory, and can be printed directly from any text editor, once the files have been unpacked.

BRB-ArrayTools Version 4.5.0 beta 2 Release

BRB-ArrayTools is a set of tools for the analysis of DNA microarray data.
BRB-ArrayTools has tools for data manipulation, such as collating and filtering data from multiple experiments, as well as tools for data analysis, such as hierarchical clustering and multidimensional scaling.
BRB-ArrayTools also annotates genes of interest by linking to NCBI databases.

System Requirements

Windows:

BRB-ArrayTools is designed to run as an add-in for Excel 2000 or later, on Windows Vista/Windows 7/Windows 8 as well as a 64-bit machine. BRB-ArrayTools is no longer supported for Excel 97/Excel 98.
BRB-ArrayTools itself will require about 200 MB of disk space, the R software about 400 MB, and the Java requires about 100 MB of disk space. It is recommended that the user have at least 512 MB of RAM to run this Software.

When installing BRB-ArrayTools or CGHTools on a 32 bit or 64-bit machine with Vista or Windows 7, please make sure you have “FULL Control” to the program files folder. (C:/Program Files) or C:/Program Files(x86)/ folder.

MS Vista/ Windows 7/Windows 8:

BRB-ArrayTools can run on MS Vista/ Windows 7/ Windows 8 and Excel 2003/Excel 2007/Excel 2010/Excel 2013.
Excel 2007, 2010 and 2013:
It is required to check Trust access to the VBA project object model.
- Click the “Office Button” located on the left-top of Excel menu,
- Click “Excel Options”, then choose “Trust center” on the left, then “Trust center settings”, then “Macro settings” on the left,
- Check “Enable All Macros”
- Check “Trust access to the VBA project object model”, and click “OK”.

Add-In:
- Click “Add-Ins” above the Trust center on the left panel.
- Click on BRB-Arraytools on the Active or Inactive applications add-ins, and then click Go on the bottom.
- Check BRB-Arraytools, BRB-Arraytools RServer, BRB-CGHTools, then click OK.

If you don’t see “Add-Ins” ribbon alongside “Home Insert . . . Review View” panel, then close Excel and restart.

If you got this “This workbook has lost its VBA project, ActiveX controls and any other programmability-related features.” Then go to this link for a fix:
http://www.asap-utilities.com/faq-questions-answers-detail.php?m=145

Additionally, for VISTA users, please make sure you have “full control” to the “ArrayTools” and “R” installation folders.

For further details, refer to the
http://linus.nci.nih.gov/download_BRBArrayTools.html

Mac:
====
BRB-ArrayTools has been tested on an Apple macbook pro machine with Windows 7 installed with Apple’s bootcamp software. The above windows system requirements holds true.

64-bit Office and 64-bit R:

This version can work on 64-bit version of Office. Additionally, the 64-bit version of R, if available, will be launched under almost all circumstances.

Installing BRB-ArrayTools and Software Components
==========

If you have Excel open, please close Excel before installing BRB-ArrayTools.

It is strongly recommended that you have administrator privileges on your machines specifically to the “ArrayTools” installation folder (typical path is C:\Program Files\ArrayTools) and the “R” folder (C:\Program Files\R).

There are three installation steps:

1) If you do not already have the Java installed, please download Java at java.com and install it.

2) If you do not already have the R software, version 3.2.2, on your computer, then you should download and execute the "R-3.2.2-win.exe" installation file from the BRB-ArrayTools download website (http://linus.nci.nih.gov/~brb/download_pre2.html), or obtain the "R-3.2.2-win.exe" file directly from the CRAN website (http://cran.r-project.org). You must install R in the default folder namely C:/Program Files/R.

3) Download and execute the "ArrayTools_v4_5_0_Beta_2.exe" installation file. If you already have a previous version of BRB-ArrayTools installed, you should install the newer version in the same installation directory as the previous version, and the newer version will overwrite the previous version. (You should avoid installing BRB-ArrayTools in a different directory than the previous version, since this would require that you go through additional procedures when loading the add-in within Excel.)

Full Installer:

The BRB-ArrayTools software download page has an option to download the Full installer. This file is a complete bundle of all the required components namely Rv3.2.2, Java, as well as ArrayToolsv4_5_0_and CGHTools.

Using BRB-ArrayTools within Excel

Once BRB-ArrayTools has been loaded as an add-in in Excel, all of its functions can be accessed from the ArrayTools menu. BRB-ArrayTools comes with a set of on-line HTML help files which can be accessed from the Help menu as well as from the dialog forms.

Changes and Bug Fixes since Last 4.5.0 Beta 1 Release Version

Analysis Tools:

1) Added DrugBank version number in the html output file generated by running the DrugBank Utility tool.
2) Fixed a bug in creating an Ingenuity IPA output file in the Class Comparison tool.

Importing, Filtering, Normalization and Annotation:

3) Modified code in GSE data importer to generate a more user-friendly Experiment Descriptors Worksheet.

Changes and Bug Fixes since Last 4.4.1 Stable Release Version

Visualization Tools:

1. Fixed a bug in the 3D Visualization of Samples tool when Excel 2013 is used.
2. Added the Ward’s linkage method in the “Dynamic Heatmap Viewer” and “Cluster Samples Only” tools.

Analysis Tools:

3. Added a new option to find differentially expressed genes by controlling the ‘Local false discovery rate’ (Efron, et al 2001) in the class comparison (between groups of arrays) tool.

4. Developed tools for differential expression analysis between two classes on RNA-Seq count data with “DESeq2” or “edgeR” packages.
5. Fixed a bug where the ingenuity IPA output file created in class comparison analysis cannot be opened in the Firefox browser.
6. Modified code in the DGIdb utility to handle one gene case.
7. Fixed an error in clustering detection parameter setting in the Preference menu.

Importing, Filtering, Normalization and Annotation:

8. Added an importer to import RNA-Seq count data.
9. Added an importer to import GSE data from GEO.
10. Modified code in the ST array importer to use the “oligo” package to allow importing of Affymetrix gene 1.0, 1.1, 2.0 and 2.1 ST arrays. The use of “aroma.affymetrix” package has been deprecated.
11. Added Arabidopsis array with the TAIRG version in the custom cdf option.
12. Fixed a bug in the “reset” function when re-filtering is applied.

Changes and Bug Fixes since Last 4.4.0 Stable Release Version

Analysis Tools:

4) Updated the random number generator package (from rsprng to rlecuyer) used in parallel computing.
5) Fixed a bug in selecting the optimal lambda when the penalized Cox regression was selected with clinical covariates in survival risk prediction analyses.
6) Fixed a bug when there were no significant genes found based on penalized Cox proportional hazards model.
7) Updated the link and version (to v4.2) for the drug bank Utility.
8) Modified code in “correlate methylation with expression” to fix an error in cases with too many missing data.
9) Modified code to allow case-insensitive match in the “Create genelist -> GO description” utility.
10) Modified code to handle cases where there are no more than 3 distinct numbers in the “column for defining a continuous response” in the quantitative trait analysis tool.

Importing, Filtering, Normalization and Annotation:

11) Modified code in ST array importer to reflect the change in the “aroma.affymetrix” package.
12) Modified code to fix an error in filtering and normalization where the Unique IDs in the gene identifiers are integers.

Changes and Bug Fixes since Last 4.4.0 Beta 2 Release Version

Visualization Tools:

1. Enhanced the Dynamic Heatmap Viewer Tool by adding a gene dendrogram (and a cut tree function), a new quantile color option, and an option dialog to allow users to change or save preferences.
2. Modified code to stop showing the “defined_genelist” column in the “gene information” table in the scatterplots.

3. Modified code to allow using gene symbol or unique id to match genes in the “highlight genes in a gene set” function in the Scatterplot Tool.
4. In the “Pairwise correlation plot” tool, the samples in the plot will be ordered based on clustering of samples with regard to correlation.
5. Fixed a bug with the “Cancel” button on the Dynamic Heatmap Viewer dialog form.

Analysis Tools:

6. Added a Utility tool to search the Drug-Gene Interaction database (DGIdb) for drug-gene interaction information on genes of interest.
7. Updated the transcription factor target gene sets using the positional weight matrices obtained from JASPAR2014.

Importing, Filtering, Normalization and Annotation:

8. Changed the default number of arrays whose expression values shown in the Filtered log intensity/Filtered log ratio worksheet from 5 to 20.
9. Fixed a bug in annotation with Bioconductor packages.
10. Modified code to use vst transformation on Illumina Expression data without the bead number column.
11. Fixed a bug in data importing with Excel 2013.

CGHTools:

12. Fixed a bug in generating the html output file for the “Identifying frequent copy number aberrations” tool.

Changes and Bug Fixes since Last 4.4.0 Beta 1 Release Version

Visualization Tools:

1) Modified code in the “Boxplot of gene expression on each array” tool for the case with large number of arrays.
2) Fixed a bug where the Dynamic Heatmap Viewer could not be launched under Windows 8.1 Operating System and Excel 2013.

Analysis Tools:

3) Modified code in the Class Comparison Tool to change the output for extreme numerical values and fix a bug in case of all survival data being censored.
4) Replaced the predicted microRNA target gene lists with the latest experimentally verified target gene lists obtained from miRTarBase database.
5) Modified code in Class Comparison tool to allow creating a tab-delimited .txt file that can be imported into Ingenuity under the circumstances of having no annotation or having missing values in fold changes.
6) Changed the html output for most analyses written in plug-in to follow html5 standard. It is to fix the problem where gene tables

may have missing borders shown on certain web browsers such as Internet Explorer.

7) Updated the Drug Bank web site link to reflect its recent change.
8) Fixed a bug in Display Data.
9) Fixed a bug in Extract Gene Expression Data where the project has only one column of Gene Identifiers.

Importing, Filtering, Normalization and Annotation:

10) Updated SOURCE annotation.
11) Modified code to allow use of justRMA() in importing Affymetrix data in CEL file format in case of more than 100 arrays.
12) Fixed a bug when empty trailing rows are present in the data file when importing Illumina Expression or Methylation data.
13) Fixed a bug in importing Illumina expression data under non-English settings.
14) Fixed a bug when missing values are present in Illumina Expression data.
15) Fixed a bug in spot filtering with unlogged dual ratio data.
16) Fixed a bug caused by redundant column names being present in Gene Identifiers.
17) Updated the download link of Affymetrix ST-array CDF.
18) Fixed a bug in downloading Affymetrix annotation packages.

CGHTools:

19) Changed the dialog box in gain/loss analysis to allow entering different threshold values for amplification, gain, loss and homozygous deletion.
20) Added more options for CBS segmentation.

Changes and Bug Fixes since Last 4.3.2 Stable Release Version

Visualization Tools:

1. A new interactive heatmap program “Dynamic Heatmap Viewer” written in C++ was created to replace current implementation in clustering genes and sample tool. The new program allows users to explore gene expression by visualizing the heatmap of data interactively through simple-to-use graphical user interface. Some of features include zoom-in of heatmap, real-time change of gene ID and array label from mouse-over gene and array, displaying sample class along with heatmap, and so on. The previously existing “Heatmap of Data” tool is deprecated.
2. Added color-coding KEGG pathway graph functionality to Class Comparison and Gene Set Analysis tools. This new tool provides a powerful visualization tool to discover up- or down-regulated genes in KEGG pathways.
3. Modified the tool of boxplot for individual genes per class under the case some genes in user’s gene set cannot be matched with gene identifiers.

Analysis Tools:

4. Developed a new module in Class Comparison and Quantitative Trait Analysis tools to automatically create Ingenuity IPA output files to be imported into Ingenuity IPA.
5. Updated Broad MSigDB collections to the latest v4.0 version. Modified Broad MSigDB download module to allow users easily update the download link themselves.
6. Separated the MIR and TFT gene sets in Broad MSigDB C3 motif gene sets.
7. Modified code to center and scale genes after hierarchical cluster analysis is performed. Centering and scaling is only conducted for the purpose of generating a heatmap.
8. Modified Fortran programs to fix an error when all samples are censored in both finding genes correlated with survival and survival gene set analyses.
9. Fixed a bug where special characters are present in data file/file names.
10. Fixed a bug in “Create Genelist with GO description” tool where the gene list was saved in ArrayTools installation folder instead of the project folder.

Importing, Filtering, Normalization and Annotation:

11. Added the option to allow importing custom Gene Identifiers files for annotation with files in mAdb format.

Installation, Registration and Support Links:

12. Modified the license key file structure to improve security.

Changes and Bug Fixes since Last 4.3.1 Stable Release Version

Visualization Tools:

1) Fixed a bug in NMF plug-in where the size of heatmap margin was not correctly assigned.

Analysis Tools:

2) Updated the KEGG pathway gene list with the KEGG.db package and modified code to change the hyperlinks of KEGG pathways to the KEGG website instead of the CGAP website.
3) Modified the name of ‘lassoed’ to ‘lasso’ in the HTML output of lasso logistic regression.
4) The LINPACK option in svd() used by survival risk prediction was removed.
5) A false warning message "Cox proportional hazards model estimation using the principal components of the full training arrays can not be completed" was removed from the HTML output in the survival risk prediction for a successful run.
6) Fixed a bug where special characters were present in certain Gene Identifier columns.
7) Fixed a bug where “Symbol.txt” did not exist in the “Annotations” folder.
8) Modified the dialog form for the “Create genelist correlated with a target gene” utility.

Importing, Filtering, Normalization and Annotation:

9) Added an option of using the “org.Xx.eg.db” package for annotation.
10) Fixed a bug where Broad C6 gene list files were displayed when user chose User Gene List in gene-subsetting.
11) Modified GEO importer to allow the user to use the annotation file downloaded from the GEO website for annotation. Also modified code in GEO importer to ask users if they want to apply log2 transformation on certain data types.
12) Fixed a VB run-time error '5' for the “Cancel” button in Data import wizard.
13) Modified C++ code to fix a hyperlink problem from the gene annotation worksheet that could cause Excel to crash.
14) Modified Illumina expression data importer to allow importing data in the absence of STDERR/STDEV columns.

Installation, Registration and Support Links:

15) Modified code to improve the encryption of the license key file.
16) Fixed a bug in activating BRB-ArrayTools where a license key did not match information in the registration data base.
17) Modified code to automatically remove KEGG genelist files for commercial or sixty-day users.

CGHTools:

18) Updated the KEGG gene list file.
19) Modified code to remove the KEGG gene list file for Commercial and sixty-day users.

Changes and Bug Fixes since Last 4.3.0 Stable Release Version

1. Upgraded R to version 3.0.1 so as to fix an error where some R packages cannot be loaded properly.
2. Fixed an error in automatic checking server for updates when clicking on the “ArrayTools” menu item.

Changes and Bug Fixes since Last 4.3.0 Beta 3 Release Version

Visualization tools:

1) Fixed a bug in 2-D Scatterplot. The 2-D Scatterplot tool will automatically re-compute up/down regulated gene index when the user changes the fold change value and the color of up/down regulated genes.
2) Fixed a bug in the heatmap of median values where the median values were out of range.
3) Fixed a bug in Clustering Genes (and Samples) -> Zoom and recolor.

Analysis Tools:

4) Fixed a bug when 'average over replicates' option was checked in the survival risk prediction analysis.
5) Modified sample size calculation tool to handle the case if sample size cannot be computed using either 50th or 75th percentiles.

6) Modified boxplot of gene expression on each array tool to work with the case if data is dual channel with individual intensities and the number of total arrays is less than 15.
7) Modified boxplot of gene expression from individual genes per class tool to show boxplots from multiple matched probesets instead of only the first matched probeset.
8) Modified plugins to handle the situation when the class variable is a mix of numerical values and empty characters.
9) Removed Table 2 in the Quantitative Trait Correlation analysis output file.
10) Fixed an error in Creating Correlated Genelist tool.
11) Fixed a bug in Quantitative Trait Correlation tool.

Importing, Filtering, Normalization and Annotation:
12) Modified code to reset the “background correction”, “average replicates” and “Common reference design” to False upon clicking the “Reset” button.
13) Fixed a bug in SOURCE annotation where the “EntrezID” column contains empty values. Also fixed a bug in SOURCE annotation when the values in a particular Gene Id column are all empty.
14) Modified RNA-Seq importer to make it more flexible.

Installation, Registration and Support Links:
15) Fixed a bug with the registration button.
16) Fixed a bug where sometimes a message “This workbook is referenced by another workbook and cannot be closed” pops up when opening Excel.
17) Updated links to message board.

CGHTools:
18) Disabled the “HaarSeg” option in Segmentation.

Changes and Bug Fixes since Last 4.3.0 Beta 2 Release Version

Visualization Tools:
1. Added the options of adding horizontal & vertical lines on the volcano plots in the class comparison dialog form.
2. Fixed a bug in Visualization of Samples where the gene expression values were not correctly passed to R to perform multi-dimensional scaling when the global test of clustering is turned on.

Analysis Tools:
3. Updated the Broad MSigDB gene sets from v3.0 to v3.1.
4. Added the Broad MSigDB C6 gene sets in the gene lists.
5. Fixed a bug in plug-ins when the experiment descriptors worksheet is sorted.

Importing, Filtering, Normalization and Annotation:
6. Updated the link for SOURCE annotation.
7. Fixed a bug in RNA-Seq importer.

8. Fixed a bug in dual-channel median print-tip normalization.
9. Fixed a problem in installing “aroma.affymetrix” package for the Affymetirx 1.0 ST-array importer.
10. Fixed a bug when a Unix .txt file could not be automatically converted to a Windows.txt file.

Installation and Registration:

11. Fixed a bug where the “Register” button in the ArrayTools activation form did not work.
12. Modified code to remove the ArrayTools update installer after the ArrayTools version is updated.

Changes and Bug Fixes since Last 4.3.0 Beta 1 Release Version

Visualization Tools:

1. Visualization of Samples 3-D plots: can be viewed through PowerPoint slideshow.
2. Scatterplot:
3. When the project has more than 32k genes that passed the filter, a non-interactive scatter plot will be shown instead of the tcl interactive image.
4. The exported gene list file name for up/down regulated genes can be changed by the user.
5. Modified code to significantly reduce the computing time for phenotype average.
6. Fixed a bug where color palettes did not show correctly with 64-bit Microsoft Office.

7. Analysis Tools:
8. Added message to show the result file location after running the “Extract gene expression data” utility.
9. Modified code to open all HTML files with the default Internet browser instead of Internet Explorer only.
10. Modified code to check the existence of Entrez ID information in the “Annotations” folder only when running gene set comparison analysis for Gene Ontology.
11. Fixed a bug in “cut tree” in Hierarchical clustering of samples alone.
12. Updated the link to the Drug Bank website.
13. Fixed a bug in ANOVA for fixed effect model when no significant genes can be found in all of main effects.
14. Fixed a bug in Lasso logistic regression when genes were not annotated.
15. Fixed a bug in Survival analysis when only one gene passed the filter.
16. Fixed a bug in Top scoring pairs analysis when the gene expression data are missing for all arrays in one class.
17. Fixed a bug in Correlating methylation with expression when no significantly correlated gene is found.

18. Increased the JAVA memory to 1024M when loading “xlsx” package so as to fix a “JAVA out of memory” error when running correlation between methylation and expression.

19. Importing:
20. Added support for importing Red channel intensity data in Agilent single channel data importer.
21. Fixed a bug where “Cancel” button does not work properly in importing Affymetrix CEL files.
22. Fixed a bug when trailing empty lines are present at the end of a horizontal data file.
23. Fixed a bug when trailing empty columns are present at the right side of a data file.

24. Installation and Registration:
25. Added a “Register” button in the ArrayTools activation form.
26. Fixed a bug where the JAVA path could not be correctly recognized in 64-bit Excel.

What's New in BRB-ArrayTools Version 4.3.0

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.

Changes and Bug Fixes since Last 4.2.1 Stable Release Version

1. 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 64-bit operating system.
2. Fixed the gene annotation problem on the HTML output in the lasso logistic regression if a subset of genes were used.
3. Fixed an imputation usage issue in the quantitative trait prediction analysis if the data contains missing values.
4. The “Match dataset against Genelist” Utility is changed to only match gene lists in the BioCarta and KEGG pathway folders.
5. Fixed a bug in computing minimum intensity filtering when all the arrays had missing values.
6. Fixed a bug in NCBI GEO importer when 64bit Excel is used.
7. Fixed a run time error 13: type mismatch in running Re-filtering in ArrayTools.
8. Fixed a bug in the Create Genelists correlated with a target gene utility.
9. Fixed a bug in Gene Set Expression Comparison when Illumina data is annotated using the bioconductor annotation packages.

CGHTools:

10. Fixed a bug when “NA”s exist in the correlation results.

Bug Fixes since Last 4.2.0 Version

1: Lasso PC plug-in: The code was modified to adapt to the latest changes in the package. Also, fixed an error that occurred when the default output folder name was modified.
2: 64-bit OS Cluster reproducibility: Re-compiled the .dll to be compatible with the 64-bit OS when running the cluster of samples.
3: Fixed an error in Single channel normalization when the reference array was not explicitly specified.
4: Modified the code to obtain relevant packages from the Bioconductor repository.

CGHTools:
1: Modified the code to handle foreign language settings.
2: Also fixed a run time error '91' in general importer.
3: Fixed a bug in writing out the HTML output file when the total number of genes is identical for all pathway gene lists.

Bug Fixes since Last 4.2.0 Beta 2 Version

Changes made to data importing:

1. Modified the Agilent Importer to make the spot size optional, so as to adapt to the new changes in the feature extraction file format.
2. Added a check when importing the data to warn the user when empty(blank) cells are detected in the unique id column.
3. Added an appropriate message to the data import wizard for Affymetrix data when the necessary detection call column is not available in the raw data files.
4. Fixed an error in reading the last column when importing the user’s specified annotation file.

Analysis tools:

1. Fixed a bug in the lassoed logistic regression when the permutation test was requested.
2. Modified adaboost R code for the syntax change in the R's package.
3. The histogram, smoothed CDF plot and the pair-wise correlation plot plug-ins now run on genes that have passed the gene filtering options.
4. Modified the code to turn off the Random Variance Model (RVM) option when the number of arrays was greater than 100 in Class Prediction, Class Comparison and Gene set Comparison tool as well as the ANOVA (fixed effect and log intensities for dual channels), random forest and adaboost plug-ins.

5. In this version, the Goeman’s test has been removed from the Gene Set Comparison tool. Also, the code now correctly reads the option related to the maximum number of genes for Gene Ontology categories.
6. Modified the code to handle the instance when a user- defined genelist file is a blank file except for the header row.
7. Modified the code in Non Negative Matrix Factorization plug-in to handle the instance when there was only a single array in one of the clusters.
8. Modifying the 3-D scatter plot code to handle the instance when the 3D graphic window could not be closed if the identify function was specified.
9. Modified the code to appropriately save the workbook after clustering of Genes and Samples was run.

CGHTools:

1. CGHTools will launch 64-bit R in batch mode when detected.
2. Modified the platform specific importer to make it more flexible for identifying Affymetrix .CNT files.
3. Removed NimbleGen arrays from CGHTools platform specific importer due to inconsistencies among versions of NimbleGen data files.
4. Removed the file extensions in CGHTools Array ID column in the Experiment descriptor Worksheet.
5. Fixed a bug in some analysis tools in CGHTools when analysis continues to run if the DOS window is not manually closed.

Bug Fixes since Last 4.2.0 Beta 1-Patch_1 Version
==========

Analysis:

1. Fixed a bug in Binary Tree class prediction when “option” button was clicked it triggered an error message “s_FilteredLogIntensity could not found in Import.txt file”.
2. Clustering genes and samples: Modified the code so that the list genes, zoom/ recolor and cut tree buttons work specifically on previously saved projects. Also, fixed an error to correctly display heatmaps for samples sizes that ranged from 10 to 40.
3. Modified the DrugBank utility to accommodate changes made to their web site.
4. Fixed the problem when the unzipping of the distributed genelist files failed on some computer systems.
5. Added a new analysis tool 'Cell type specific significance analysis of microarrays (csSAM)' under the plug-in menu option.
6. Predict quantitative trait analysis was modified to work with the latest “lars” R package.
7. almostRMA was modified to check for the consistency of chip types when importing. An informative message will be shown if different chip types are detected in the same folder.

8. Class prediction was modified to fix an error in creating HTML output for a single significant gene situation.
9. Random forest analysis was fixed so that users do not have to specify class labels for test cases (predict arrays).
10. Adaboost analysis has been fixed for an error that occurred when the random variance model was selected.
11. Volcano plots in the class comparison output will use 1e-7 as a threshold for genes with p-values < 1e-7. Parallel coordinate plot was modified to fix a problem when the random variance model was selected. Also, added information about getting the interactive feature to work in IE with the ActiveX component.
12. PAM was modified to fix an error caused by using -9999999 as the missing value. Also, corrected the HTML output to show the filtering parameters.
13. R package Cairo was modified to use http://www.rforge.net as the repository to download from instead of http://cran.r-project.org due to a bug in the binary package from CRAN web site.
14. The HTML output in SAM analysis has been modified to now display the reason when the program has run into a memory issue due to a very large number of permutations specified.
15. Fixed a bug in Affymetrix quality control when the input file names had a “.”.
16. Time course heatmap has been enhanced to include the order to the gene list table output so users can sort genes based on the heatmap ordering.
17. Fixed a bug for box plot plug-in to correctly apply log transformation to the data before normalization.
18. The PowerPoint created from MDS tool now launches in Office 2010.

Importing, Filtering, Normalization and Annotations:

19. Fixed a bug in lowess normalization when the intensity filter is turned off, the software still performed the intensity filtering.
20. Fixed a bug in gene sub-setting that was caused when a gene list file was missing but the program tried to load it.
21. Modified the code to handle instances where the check box was turned off in filtering dialogs but the corresponding text box was empty.
22. Fixed the bug in the ST array importer when the cel file names contained “.”
23. Fixed a bug in importing to clean out old files in the same project folder when overwriting a project folder.
24. Fixed a bug in importing that would not create a project specifically when the raw data folder was directly under root directory.

25. Modified the GenePix Single channel data importer to handle new format of data.
26. Illumina importer: Modified code if the input data did not have the “beadNum” or the “Detection” column, then a log2 transformation will be used. Also added a message that Probe IDs will be converted to NuIDs when the data is annotated through Bioconductor. Fixed an error when there are trailing spaces in array names.
27. Fixed a bug specific to single channel data that was caused when the average replicate option was selected.
28. Modified the code to allow users to have the option of not applying background adjustment even if background column is available.
29. Added support in SOURCE annotation for canine species.

ArrayTools:

30. This version is now compatible with Excel 2010.
31. A permission problem for the R package installation on Windows Vista/7 has been addressed. A writeable directory will automatically be selected to install R packages.
32. Added code to automatically save the current project when users re-filters or annotate the data.

CGHTools

33. Fixed a bug in Bug tracking tool in CGHTools where the actual error message was not recorded.
34. Fixed a bug in importing CGH Illumina data where Log.R.Ratio instead of “Log R Ratio” is used in column header.
35. Fixed an indexing error in Pathway analysis in CGHTools when MAD is selected for gain/loss determination.
36. Added support for Canine genome build in CGHTools (development).
37. Fixed an error in CGHTools when parameter passing to R in foreign regional language setting.
38. Modified the CGHTools code to handle HaarSeg package installation problem as the hosted package server does not support R 2.12.x anymore.
39. Fixed a bug when the chromosome column contained “NA”s in the Chromosome information file.

Bug Fixes since Last 4.2.0 Beta 1 Version

1) Fixed a run-time error that occurred on re-filtering converted projects.
2) Fixed an error during project conversion, when gene sub-setting by gene identifiers was run in the previous version of the project.
3) Fixed an error in dual channel lowess and print-tip lowess normalization, when background correction was not applied.

4) Fixed an error in single channel housekeeping normalization, when sometimes the browse button for the house keeping file was not get activated.
5) Fixed an internet connection error and file-writing error when creating the license key file.
6) Added the Clustering ordered sample id list to the output in the Clustering genes and samples.
7) Fixed a bug in Quantitative trait analysis tool for single channel data.
8) Modified the code to handle a foreign language error caused when re-filtering.
9) Fixed an error in class comparison when no significant genes were found.

What's New in BRB-ArrayTools Version 4.2.0 and CGHToolsv1.2.1

Visualization tools

New rotating 3-D interactive plot of samples. Axes are user selected Biocarta/Kegg pathways, gene lists or individual gene symbols. This 3-D plot can now be saved and launched in MS PowerPoint.

Data Import

Re-organized the code related to importing. In this version, averaging replicate spots, background subtraction and common reference design are now part of the filtering options. The importing of Affymetrix multi-chip sets is not supported in this version.

Affymetrix .CEL files

Custom Chip Definition Files (CDF) from the University of Michigan can be used when importing .CEL files.

Annotations:

Additional support has been added to SOURCE annotations for Agilent, Affymetrix and Illumina data.

Normalization

For single channel data two new methods have been added. The quantile normalization method and an option to normalize each array based on a specified percentile and target intensity.

Analysis Tools

Lassoed logistic regression plug-in:
This plug-in implements Friedman et al (2008)’s method to fit a logistic regression model to predict a binary class variable using gene expression values and optional standard clinical covariates. It uses a L1 penalized maximum likelihood method and performs complete cross-validation evaluating prediction accuracy of genomic model to clinical model to combined model.

Class comparison: For this release, the interactive volcano/parallel co-ordinate plot is included in the HTML output.

Heatmaps:
Added an option to scale for single channel data in clustering of genes and samples.

CGHTools:
Added an automatic bug reporting tool for the various analysis.

Changes and bug fixes since v4.1.0 Beta_2 Release
==========
1: Class Prediction: Fixed an indexing error in the HTML output table for the prediction of new samples in cases where the true class label is available.
2: Mixed Effects ANOVA: Added an option to permit an additional fixed effects to the model.
3: Fixed an error when no genes were found in survival risk prediction.
4: Fixed an annotation error when the user selected to annotate the project with their own gene ids. Also, modified the code to handle the case where the project was saved on a different drive than where ArrayTools was installed.
5: The global test option for MDS now runs.
6: GEO importer now handles an additional data type called expression profiling by arrays.
7: Users are now permitted to use non-integer values for spot size filter.
8: Modified the R code to correctly read the array ids with trailing spaces for median normalization in single channel data.
9: Enhanced the dialog in the extract gene expression data plug-in.
10: Fixed a bug in generating analysis related heatmap for paired data.
11: Modified the code to handle changes made to the BROAD institute's gene signature databases.
12: Modified the code to support the ArrayTools automatic updating for VISTA and Windows 7 users.

Enhancements and Bug Fixes since Last 4.1.0 Beta 1 Version:

1: Added a new option to filter genes for single channel based on minimum intensity.
2: Enhanced the 2-D and 3-D scatter plot tools.
3: Fixed a critical error in ST Array importer that affected the normalized log intensity values.
4: Re-compiled the Fortran program for almostRMA to fix a dll problem for windows 7 users.
5: Added gene names to the heatmap zoom/recolor option in clustering.

6: Fixed a time series error that occurred in the heatmap when there was only one array per time point.
7: Fixed a minor error in the RVM when the variance for a give gene was zero.
8: Fixed an error in Affy. cel file importer for the MAS5.0 option, to correctly run the detection call filter in spot filtering.

What's New in BRB-ArrayTools Version 4.1.0

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.

Changes and bug fixes since v3.8.0 stable Release:
==========

1: Geneset comparison tool- Fixed an indexing error in the Fortran program related to allocating common block variables for the Random Variance Model estimation.

2: Handling redundant probes within gene set comparison tool- An indexing error has been fixed when the redundant probe option was selected. Also,the Random variance model uses a filtered list of genes as opposed to the reduced list of genes based on redundant probes.
3: Data Import Wizard- The code has now been modified to correctly read the spot flag string for the Agilent importer.
4: Genelists- Modified the genelists files that were included as part of the distribution to remove corrupted files in the transcription factor and PFAM protein domain gene sets for mouse.
5: Class comparison- Fixed a problem related to the parallel coordinate plots in specific data sets with missing values and using a blocking variable.
6: Fixed a bug in the survival analysis where the survival curve was not be shown if the gene expression data was too skewed.
7: Quantitative Trait Analysis- When requested the HTML output now shows the permutation p-values.
8: ANOVA plug-in log intensities- The HTML output now displays the geometric means.
10: Added support for annotating with Bioconductor Xenopus laevis and Xenopus tropicalis.

CGHTools:
1: In this version, the genome build information is accurately saved when the user selects to specify a chromosome file.

Changes and Bug fixes since the last 3.8.0 Beta_3 Version:

1: Modified the code for various analysis tools to be compatible with the latest Rv2.10.
2: The SOURCE annotations has been modified to accommodate changes made to species names by Stanford.
3: Fixed an error that occurred in specific cases, related to Survival analysis tools when the status for all the arrays is 1 (1= death).
4: The Fortran code was modified to increase the precision of test statistic values from SAM analysis. Also, removed a redundant imputation step that was previously performed on paired data.

CGHTools:
1:Modified the R code to use random sampling method (n=1000) instead of normal approximation to obtain the null distribution of the statistic for the pathway enrichment analysis.

Changes and Bug Fixes since the last 3.8.0 Beta 2 Version:

ArrayTools:
1: GenePix importer: Fixed a bug related to the Filtering and normalization options specified at the import step were turned off.

2: Lowess Normalization: The spot filter was not applied to the individual intensities but only to the log ratio data when computing the Lowess smoother function.
3: Modified the code to read the print tip block variable when importing the data.
4: The spot filtering is available when importing Affymetrix .CEL files with the MAS5.0 option.
5: Fixed the scattterplot experiment vs experiment for individual log intensities to display the data values.
6: Updated the web link related to downloading the BROAD institute's genesets.

CGHTools:
1:The output for the GISTIC has been corrected to show the Benjamini-Hochberg estimated false discovery rate rather than the Family wise error rate. Also, modified the code to use resampling method instead of normal approximation to find the null distribution of Gistic statistic (B=10000).

2: Fixed an indexing error related to the MAD factor calculation to now include the sex chromosome. This could affect GISTIC, Correlation and Pathway analysis if MAD method was selected.

Changes and Bug Fixes since Last 3.8.0 Beta 1 Version:

1: Fixed the bug related to HTML error caused when Gene ontology observed vs expected analysis for class comparison was selected. The Fortran code has been corrected to handle an indexing problem. Also, fixed an error in the Volcano plot that occurred in some instances when the Fortran program wasn’t completed but the plot was launched. The code has been modified to appropriately handle the situation when fold change option was selected with univariate permutation. The fold change option is not permitted when the blocking variable is selected.
2: Gene Set Comparison: Modified the code so that the ArrayTools path is no longer hard coded and also adjusted the heatmap plot dimensions.
3: Modified the code for Gene set comparison to display the results when the GSA package failed due to the limit of a total of 110 unique genes.
4: Survival analysis: The code has been modified for the case when no genes were found significant. Corrected the gene list file created from survival analysis.
5: Modified the R code in different tools, to handle the appropriate messages that were previously displayed using the windialog() function. The R code has also been modified to handle the latest impute package developed under Rv2.9.0
6: BROAD web server: Modified the code to use the http link instead of the ftp link based on changes made by the Broad institute.
7: SAM: Modified the Fortran code to handle the situation when the data had too many missing values then the corresponding CDF for the F distribution had negative values. Also, the HTML output has been modified to represent a consistent plot for the positive and negative significant genes.

8: Modified the LARS plug in to use a different random seed. The HTML output has been enhanced to include a table of actual and predicted responses used in the scatter plot. The formula for predicting a new sample has been corrected.
9: Illumina importer: The software now correctly reads files with the underscore character and can import ENTREZ ID as well. The code has been modified to import Target ID when Probe ID is not available.
10: Modified the code for SOURCE annotation to allow users to select ENTREZ ID as one of the identifier to download the annotations.
11: Fixed an erroneous option in the gene identifiers import dialog.
12: Housekeeping gene normalization: Fixed an error in data sets with > 65k rows. Also, maximum number of housekeeping genes in the genelist file has been modified to be larger than 3000 rows.
13: The cancel button when creating the project workbook or now cleans up the appropriate files.
14: Modified the VBA code to use http instead of ftp for updating ArrayTools from the linus server.

CGHTools:
1: Fixed a bug caused when the sample ids had numeric values.
2: Made minor changes in the code to appropriately prompt the user to open a CGH project on clicking different menu options.
3: Added information to the HTML on the gain and loss thresholds used Pathway, GISTIC and correlation analysis.

What's New in BRB-ArrayTools Version 3.8.0

1:Enhanced class comparison: The output of class comparison between groups of arrays includes a heat map of the significant genes as well volcano plots (for 2 class levels)/parallel coordinate plots (more than 2 class levels). An option to restrict the genes by Fold threshold is also implemented.
2:Gene Set Comparison: Added an option to handle redundant probes that correspond to the same gene. Enhanced the output to provide heat maps for the significant gene sets.
3: Time course analysis: Enhanced the output by providing a heat map for significant genes.
4: NMF plug-in: A new clustering method using non negative matrix factorization method has been included as a plug-in tool in this release.
5: Least Angle Regression (LARS): Implemented a new tool for prediction of a continuous response variable.
6: Utility: Added an option to automatically download packages from CRAN/Bioconductor that are needed for analysis. This will permit the user to perform various analyses even when not connected to the internet.
7: Importer: Added an option to import and normalize Illumina data using the ‘lumi’ package.

8: This version has the capability to download annotations for additional species from Bioconductor.
9: The ANOVA plug-in has been modified such that Table 2 for the fixed and mixed effects model has been removed to simplify the output.
10: Modified the installer so as not to register the shdocvw.dll due to a windows security update which no longer has the permission.
11:Fixed a VBA inconsistency when picking the median array for even number of arrays to always pick the left array.

What's New in CGHTools Version 1.1.0

1: Added an option to import inferred copy number data using the general importer.
2: Individual HTML outputs are generated for Segmentation and gain/loss analysis in this release.
3: Gain/Loss Analysis: Added options for user to determine gain and loss based on arbitrary segmentation log ratios or the MAD factor multiplied by the segmentation mean log ratios as well enhanced the output by adding frequency plots.
4: Implemented the GISTIC tool to systemically identify regions with frequent and significant copy number aberration.
5: Capability to assign summarized values on unique gene symbols for each array based on the inferred integer copy numbers or segmentation data.
6: Implemented a pathway enrichment analysis tool using this gene data.
7: Added an option to create an expression project(BRB-ArrayTools project) from the gene data such that further expression analyses can be performed using ArrayTools.
8: Added a feature to integrate gene expression project data with CGH data by performing a correlation analysis.
9: Included a sample data set with the distribution.

Bug fixes since the last 3.7.0-Patch_1 release:

1: Source website recently modified their link for downloading files in the batch mode and this has caused an error when trying to annotate the data from Source. The code has been modified to reflect the new web page link.
2: Some Excel 2007 users have reported a problem after the collation is done but when writing the gene identifiers. The error appeared to be caused by the Excel built-in worksheet copy function not working properly. This has been fixed by copying a range of cells instead of the entire worksheet.
3: Added a message to run the “Cut Tree” function to obtain the cluster reproducibility measures.
4: Also, fixed an error when annotating from SOURCE if the gene identifier had zeros.
5: Modified the Fortran program for top scoring pair plug-in to better handle large data sets.

Changes and bug fixes since the last 3.7.0 stable Release version:

1: rscproxy package for Rv2.8.0 now gets installed from ArrayTools instead of CRAN.

2: Fixed an error in Clustering genes and samples for the gene subset option.
3: Source Annotations: The program now recognizes both GeneId and LLIDs from Source annotations.
4: ScatterPlot Phenotype averages: The utility to download genelists now works.

Changes and bug fixes since the last 3.7.0 Beta_2 version:

1: Fixed a run time error in the ‘zoom and re-color’ button for clustering of genes and samples.
2: Fixed an error in single channel normalization using median across groups of arrays.
3: The utility to download a genelist to a file now works for scatterplot experiment vs experiment tool.
4: Corrected the redundant message that appeared when Lowess normalization is selected.
5: Updated the code for source annotation, as SOURCE has now replaced locuslink Id with geneID.
6: Modified the Fortran code for class comparison to use 100,000 as the number of permutations in the approximation method. The HTML output now correctly reflects p-value < .00001 instead of <.0000001 if the permutation p-value is zero.
7: Modified code to run annotations on yeast and download corresponding Gene ontology data using the Bioconductor packages.
8: Fixed an error in the quantitative trait analysis tool that only happened if there was a perfect correlation among some of the genes.
9: Fixed an error in the Fortran code of the class comparison tool that occurred in certain data sets. The program failed for the randomized block design when there was missing data.
10: Fixed an error in cluster that was caused when the worksheet name was not appropriately updated.
11: Modified the R code so as to be compatible with latest Rv2.8.0 release.

Changes and Bug Fixes Since Last 3.7.0 Beta 1 Version:

1: Modified the installer to correctly install CGHTools.
2: Fixed a bug that affects single channel data if median normalization is selected and the user specified a reference array then spot filtering was not applied to the reference array.
3: Fixed an error in averaging replicate spots to correctly apply the spot filtering on all the spots. The error was caused by VBA code initializing the first value to 0.
4: Fixed an error in the data import step that incorrectly imported dual channel ratio data.

5: Fixed an error in single channel normalization by groups that affected non-contiguous groups.
6: Fixed an error when importing Agilent's dual channel intensities.
7: Fixed a bug in Class Prediction for the recursive feature elimination method when there was missing data.
8: The output in Binary Tree prediction now correctly displays the Geometric means.

What's New in BRB-ArrayTools Version 3.7.0

The new version introduces many new features. A new tool called CGHTools is shipped with the same installer of BRB-ArrayTools. The CGHTools is used for the analysis of array Comparative Genomic Hybridization data.

1. ANOVA plug-in on log intensities: Added the pairwise contrast analysis option.

2. Plugin: Sample size estimator for 2 classes

3. Gene Set Comparison: Added Efron-Tibshirani’s GSA maxmean test and Goeman's global test. The structure of the HTML output is simplified to give a comparison of all tests. The Hotelling test was dropped.

4. BROAD gene set collections: Updated the Broad Institute Molecular Signature Database (MSigDB) including positional, curated, motif, and computed gene sets. Added 'rat' and mouse species for the curated gene sets group.

5. Class prediction: Added the ROC curve to the HTML output for BCCP predictor.

6. ScatterPlot: Added a button to export the genelist. Also, modified the name of the scatterplot for phenotype averages to reflect the class variable.

7. Clustering Heatmap: Enhanced the heatmap to provide a color gradient option that the user can select from a color pallet. Also, modified the heatmap zoom-in feature to allow the user to specify the gene identifier to be displayed.

8. Normalization for Single Channel: Modified the code to allow the user to normalize the data by groups of arrays.

9. GO download: Modified the code to obtain the Gene Ontology files from Bioconductor.

10. Affymetrix annotations: Modified the code to download the Affy annotations from Bioconductor.

11. DrugBank Link: Added to the "info" link in the HTML output, a link to query using gene symbol.

12. Updated the import wizard: The automatic importers for Affy, Agilent, Gene Pix, and mAdb can now be accessed using the data import wizard.

Changes and Bug Fixes Since Last 3.6.0 Stable Version:

1: Class Prediction: Fixed an error in the Bayesian compound covariate predictor and the threshold in the prediction rule of the Compound covariate predictor. For single-channel, data will not be median centered gene-by-gene.
2: Survival Risk Prediction: The prediction rules for genes only and combined models are displayed in HTML output.
3: ANOVA-based plug-ins and time course plug-in: Output gene lists for p-valued and FDR thresholds separately.
4: Data Import: Fixed a bug when importing dual-channel ratio data with the data import wizard. Note it has no impact on importing dual-channel ratio data with the general format importer.
5: Data Import: Fixed a bug when importing two-color Agilent data with the data import wizard. Previously red channel is taken as the reference channel. Now it is corrected to use green channel as the reference.

Changes and Bug Fixes Since Last 3.6.0 Beta 3 Version:

1: Class Prediction: Fixed a bug in the class prediction output where the t-statistic column had 1e-07 values instead of negative values. Also, for the CCP and DLDA prediction methods modified the code to handle missing values when computing the weights and threshold.
2: RVM: Increased the limit on the number of genes to 500K as well as increased the corresponding stack size.
3:Quantitative Trait analysis: Fixed a bug where the HTML output showed 1e-07 instead of negative values for correlation coefficients.
4: Zoom and recolor clustering: Previously, the class column selected to label the experiments was not displayed but this has been fixed in this release.
5: Survival gene set comparison: Fixed an error that was caused when incorrectly loading the default parameter file.
6: Modified the gene index in the Fortran code to handle more than 9 digits in various analyses.
7: Survival Risk Prediction: Modified the tool for the special case when no genes are selected in the combined model such that when cross-validating the model the gene with the smallest p-value together with clinical covariates will be used in the Cox regression and prediction.

Changes and Bug Fixes Since Last 3.6.0 Beta 2 Version:

1:Gene Set Comparison: Added a new family of gene sets from Pfam and SMART Protein Domain.

2: Class Comparison: Fixed an error that occurred when the p-value for the global test option was selected for class comparison analyses.
3: almostRMA: Fixed an error in launching the Fortran program for the almostRMA method.
4: Gene Set comparison-User defined gene list: Modified the code to correctly match the gene identifiers specified when using the gene list comparison option.

Changes and Bug Fixes Since Last 3.6.0 Beta 1 Version:

1: GEO Importer: Modified the code to allow users to save and unzip files under the desktop directory.
2: Data Import Wizard: Corrected a warning message that occurred when matching the unique ids with the gene identifiers file during collation.
3: Random Variance Model: Modified the Fortran code to correctly handle large a or b values which occurred when the RVM assumption was not met.
4: Class Prediction using Recursive Feature Elimination: Modified the code to exclude a gene that had all missing values within a specific class.
5: ScatterPlot: Fixed an error that occurred when running the Gene subset option with the scatter plot tool.
6:Affymetrix data: Modified the code to include the gene symbol and description in the gene identifiers worksheet and binary files after the data was annotated using Affy annotations.
7: Hotelling’s T-square test for paired data: Modified the code to correctly use the paired data when running the Gene set expression comparison tool with Hotelling’s T-square test statistic.
8: ANOVA of log intensities plug-in: Added to the HTML output, the geometric mean intensities for each class.
9: almostRMA: Enhanced the tool by replacing the R code with Fortran to significantly reduce the execution time.
10: Class Comparison: Modified the code to allow the analyses to be performed with a minimum of 2 arrays per class.
11: KEGG Pathways: Updated the pathways to reflect the discrepancy in pathway data file for hsa04110.
12: Class Comparison-blocking factor: Fixed an error caused due to missing values when a blocking variable was used in Class comparison.

Additionally, this version of BRB-ArrayTools is compatible with Excel 2007. Please refer to the ReadMe.txt file located under the “ArrayTools” installation folder for more details.

What's New in BRB-ArrayTools Version 3.6.0

The system architecture has been modified in this version of BRB-Arraytools to handle more than the Excel limit of 65,000 rows. The gene identifier and gene annotation information in now stored binary files.

This version of BRB-ArrayTools is compatible with MS Vista and Excel 2003.

Data Import:

1)GEO importer: This tool allows users to automatically import a GDS dataset from the NCBI Gene Expression Omnibus (GEO) database into BRB-ArrayTools.

2) Agilent importer: The data import wizard now automatically recognizes the format for dual channel Agilent data and directly imports the background subtracted intensities and annotations.

3)Affymetrix .CEL files: (i) For large number of .CEL files (greater than 100), to avoid memory problems, we have implemented a new method called 'almostRMA'. This method uses a subset of arrays to compute the quantile normalization and probe effects model and then applies these to all the arrays in the data set. (ii) A new option to compute MAS5.0 probe set summaries from .CEL files has been included.

Analysis Tools:

1)Gene Set Expression Comparison: We created two new families of gene sets that can be used within the Gene Set Expression Comparison tool. One family contains the set of genes that are targets of a transcription factor; one set for each TF, with the option to use experimentally verified targets or computationally determined putative targets. The second family contains a set of computationally determined putative targets for each microRNA.

2) Survival Gene set Expression Analysis: This analysis tool finds sets of genes for which the expression levels are correlated to survival. Similar to the Gene Set Expression comparison tool, this tool can be used to analyze Gene Ontology categories, Pathways, micro RNA targets, transcription factor targets and user defined gene lists.

3) Enhanced plug-in ANOVA of log intensities: This enhanced plug-in replaces the Class comparison tool between Red and Green channels. The plug-in is used for finding genes differentially expressed between two classes for two-color arrays without a common reference sample. It can also be used to compare samples of one class with the reference samples in the common reference design.

4) Class Prediction: We have implemented a new option for gene selection based on recursive feature elimination. The user specifies the number of genes to include. Starting with a full model the method excludes genes whose correlation with outcome is minimal. This reduction continues until the target number of genes is reached. The recursive feature elimination is applied from scratch within each cross-validated training set. Although recursive feature elimination is based on a support vector machine model, any type of classifier can be used for the genes selected for the training set.

5) Bayesian compound covariate predictor: We added an option of not predicting any class if the greatest posterior probability does not exceed a user-specified threshold. The HTML output now also displays the predicted probability.

We provide a new utility to create and save for further analysis a list of genes that are correlated to a user-specified gene based on a user-specified threshold.

We modified the format of the genelists that get generated from an analysis tool to include gene annotation information whenever available. This facilitates use of such gene lists with data from different projects or with different platforms.

This version has the capability to simultaneously run more than one analysis tool within a project.

==========
Bug fixes since v3.5.0-Patch_1 Release:
==========
1) Dye Swap: Using the data import wizard or the general format importer, fixed a bug to correctly compute the log ratios for the dye swap arrays.
2) Average over replicate spots: Fixed a bug in the average over replicate spots that occurred when using the data import wizard or the general format importer.

Bug fixes since V3.5.0 stable Release:
==========
1) 0.632+ bootstrap: Fixed an error caused due to incorrect dimensioning of a variable.
2) Cross-validation: Class prediction now correctly labels unclassified samples as NA instead of NO.
3) Clustering Fixed a run-time error caused due to a missing temporary worksheet.
4) Data Import Wizard: Modified the code for a more stringent string match to identify Affy data.

Changes and Bug fixes since the last 3.5.0-Beta2 Version:
==========
1) Gene Set Expression Comparison: Fixed a bug that occurred when the Random Variance Model option was selected; it was not used in the analysis of GO categories and Pathways.

2) False Discovery Rate(FDR): The False discovery rate reported in the HTML output has been corrected. The magnitude of difference to the previously reported FDR values appears small (e.g 10^(-2)).

3) Broad/MIT Pathways: Modified the code to accommodate for the changes made on the Broad/MIT web page. Enhanced the HTML output by providing hyper-links for some of the gene sets.

Changes and Bug Fixes Since Last 3.5.0 Beta 1 Version:
==========
1)Data Import Wizard: Fixed the run time errors caused due to long file paths and file permission.
2)Average replicate spots: Modified the new data import wizard to now correctly pass this option.
3)Class Prediction: Fixed the error occurred when the Bayesian Compound Covariate predictor was selected but the compound covariate predictor was not selected.
4)SAM: Modified the precision for the fold difference variable in Fortran code to handle large values.
5)Survival Risk Prediction: Fixed an error in which the prediction model did not include the covariates when fitting the 3rd model (model of Clinical covariates and gene expression).
6)Rv2.4.0: Modified various R functions in the code to be compatible with Rv2.4.0

==========
What's New in BRB-ArrayTools Version 3.5.0 Release
==========
Data Import Wizard
A new data import wizard assists users in importing their data into BRB-ArrayTools.

GC-RMA
The GC-RMA method for computing probe set summaries from Affymetrix .CEL files has been implemented.

Analysis Wizard
A new analysis wizard guides users in selecting the appropriate analysis tools for their research question and experimental design.

Survival Risk Prediction:
Enhanced to allow up to 3 risk groups and 3 clinical covariates.

Class Prediction:
A new method called the ‘Bayesian Compound Covariate predictor’ has been included for two classes. It provides a predicted probability of class membership for each class and a threshold for withholding prediction.

The Top Scoring Pair class prediction plug-in has been extended to use multiple pairs of “synergistic” genes. For the greedy pairs option we have enhanced the output to include the gene pair information.

0.632+ bootstrap Cross-validation:
The 0.632+ bootstrap method of “cross validation” replaces the 0.632 method for estimating prediction error.

Gene Set Expression Comparison

We have added a method for testing whether a pre-defined gene set contains genes that are differentially expressed among specified classes. The method is based on testing whether the top principal components of the genes in the set are differentially expressed. The multivariate Hotelling's T square test is used (Kong et al. Bioinformatics 22:2373, 2006).

Affymetric Quality Control Plots for. CEL files
We have added a utility to provide quality control plots and RNA degradation plots for projects imported using Affymetrix CEL files.

Clustering:
We have improved the color scale for the heatmap in BRB-ArrayTools. We have also added an option to median center single channel data when using the Cluster 3.0/Treeview tools.

Preferences:
Added a preference menu option to allow users to modify certain preference parameters for BRB-ArrayTools.

Log File:
A log file has been added which records the parameter options used at data importing and analysis.

Mac Users:
This version has been successfully tested with Windows XP professional running on Apple macbook pro machine. The windows XP professional was installed with Apple's bootcamp software.

This version of BRB-ArrayTools can be downloaded from
http://linus.nci.nih.gov/~brb/download.htm

==========
Changes and Bug Fixes Since Last 3.4.0 Beta 2 Version:
==========
1) Random Variance Model: Modified the code in the Random Variance model estimation to handle missing values consistently in Class Comparison, Class prediction and ANOVA plug-in tools.
2) Time Series Plug-in: Modified the plug-in so that the 'time' variable is a numerical value instead of a factor. Additionally, modified the model (C) to include the interaction between class and time**2. Significant genes for the interaction terms in model (C) won't be fitted to the model (B) where the interaction terms are not included.
3) Class Prediction: Fixed the SVM error message that shows up in DOS windows when the optimization process did not terminate with a limit of 99999 iterations.

4) Exact Number of Permutations: Fixed the error to correctly use the exact number of permutations for the multivariate permutation tests. Previously, this was always set to false. This bug fix has been implemented in the Class Comparison tools, Survival Analysis and Quantitative trait tool.
5) Quantitative Trait Analysis: Corrected the HTML output by removing Global test p-value from the HTML output.
6) Scatter Plot: Fixed the flashing of scatter plots when selecting/deselecting multiple points. Also fixed an error that occurred in the experiment vs. experiment plot, when spot flag range filter or spot size filter used non-integer threshold values.
7) Data Import using the horizontally aligned file format: Fixed the run time error regarding header line and first data line limit being 2048 char in the drop down boxes
8) Gene Subset: Fixed the gene subset selection using genelist with GenBank accession ("GB acc") type of identifiers.
9) "Click to display the data": Fixed an error on the "Filtered log ratio/intensity" worksheet so that if a numeric sort column is selected, then a numeric sort will be performed rather than alphanumeric.
10) Gene set expression comparison: The output genes for significant genesets are now correctly written to "Genelists" folder. Previously, the names of the significant genesets had been output to the “Genelists” folder.
11) Non-English Language Users: Implemented a bug fix for non-English language users to check if the decimal point (.) is being correctly passed instead of the comma (,) for some parameters in various analysis tools.
12) Users’ Manual: Updated User's Manual sections on NCI mAdb collation, GenePix collation, and format of user-defined genelist files.

Improvements and Bug Fixes in Version 3.4.0-Beta-2:
==========
1: Collation: Averaging duplicate spots:
Fixed a bug in averaging duplicate spots. When an array contained more than 10 replicate spots the bug prevented the averaging of some spots. This is a problem for GenePix files because spots with "blank" or "spot id" were considered replicated and consequently averaging was not properly done for the subsequent spots.

2: Horizontally aligned File Format:
This version of BRB-ArrayTools can now collate data for more than 248 arrays using the horizontally aligned file format.

3: CEL File Import Wizard:
An option has been added to create an experiment descriptor file template when collating .CEL files.

4: GenePix File Import Wizard:

Added an option to specify if any experiments are Reverse Fluor.

5: Filtering:
Fixed a filtering bug for Single Channel data, to turn off the "Percent Absent" filter if the data did not contain the Detection call.

6:Normalization:
Fixed a type mismatch error in Single Channel data when median normalization was selected.

7: ScatterPlot: Phenotype Averages
Fixed the runtime error, which occurred when there were missing values in the data and the phenotype, had more than 3 levels.

8: Clustering -Samples:
Fixed the printing of dendrogram labels for more than 256 arrays. Additionally, moved clustering of samples to separate 'Cluster samples' sheet instead of 'Cluster viewer'. Added a new feature to dump dendrogram labels automatically to text file Fixed a bug in which the median SD previously could not be computed from the data for when the Cluster reproducibility option was selected.

9: Clustering –Genes and Samples:
Fixed the "Zoom and Recolor" button, as previously, this button would not work outside the same session in which Clustering was performed. Fixed a bug where the array labels were misnamed or missing in drop-down boxes, when the experiment descriptor chosen for labeling did not contain unique labels. Modified the zoom and recolor dialog so that the color scheme matches the original color scheme, rather than always resetting to multicolor/quantile.

10: Survival Risk Prediction:
The output now contains the list of significant genes as well as the coefficients of the supervised principal components for the regression model. Fixed a bug when the "use separate test" option was selected and an array was labeled as "exclude". The K-Fold CV option is now enabled. Additionally, added to the HTML output the percent of variability explained by the principal components and the correlation between the significant genes and principal components.

11: Class Comparison:
Removed the p-value for the Global test when the univariate significance threshold option is selected.

12: Class Prediction:
Fixed an error in the Compound Covariate Predictor method which occurred only when using K-Fold or .632 Bootstrap cross validation options and the data contained missing values.

13: Downloading annotations from SOURCE

Fixed the bug for downloading Gene annotations from the SOURCE website when opening a previously collated project for which the data was not annotated.

14: Fixed the string match for gene symbols in the gene subset selection and annotations to be case insensitive.

15: SAM:
Modified the code so that the redundant error message in the DOS window will not occur when no significant genes were found.

16:PAM:
Can now handle an output folder name other than the default.

17: Plugins: Top Scoring Pairs
The plugin has been extended to allow for k gene pairs.

==========

What's New in Version 3.4
==========
We have re-designed the architecture of BRB-ArrayTools so that there is no longer any restriction on the number of arrays that a project can contain. The expression data is no longer saved as an Excel worksheet and so we are no longer limited by Excel’s restriction on the number of columns in a worksheet. We have tested the system with up to 1000 arrays per project. For large numbers of arrays, you need lots of random access memory, but that is relatively inexpensive. We have provided a utility that enables you to view the expression data (up to 100 arrays at a time) if you wish. Projects collated on previous versions of BRB-ArrayTools will automatically update to the revised format when the project is opened in version 3.4.

The architectural changes also speed up the analyses by passing data to R only once. This speed up is particularly noticeable for the analysis of large projects.

1) Survival Risk Group Prediction.
Version 3.4 now contains a tool to provide a multi-gene predictor of survival risk group. This is done without discretizing the survival data.

2) Gene Set Expression Comparison Using Broad/Whitehead Signatures and Pathways
We have now consolidated GO, Pathway Analysis and Gene list Comparison tools into a single tool called GeneSetExpression Comparison. It is now enabled to apply to the signatures and pathways contained in the Broad/Whitehead database of signatures. Version 3.4 contains a link to the Broad/Whitehead website and facilitates easy downloading of the requisite data and integration into BRB-ArrayTools.

3) Create User Defined Gene List Based on GO Terms

We have provided a utility for the user to create a gene list containing genes whose Gene Ontology annotations contain any of a set of user-specified character strings. Such user created gene lists can then be used to restrict any of the BRB-ArrayTools analyses.

4) Top Scoring Pairs Class Prediction

Version 3.4 provides a plug-in that implements the “top scoring pair” class prediction algorithm published by D Geman and his co-workers (e.g. D Geman et al. Statistical Applications in Genetics & Molecular Biology 3, 2004; L Xu et al. Bioinformatics 21:3905-11, 2005; AC Tan et al. Bioinformatics 21;3896-3904, 2005). We have implemented this algorithm as a plug-in. It can be easily run from the BRB-ArrayTools plug-in sub-menu and its output is very similar in format to that of the usual class prediction tool.

5) Improvement of User Dialogs

We have made changes in the user dialog pages for several analysis tools to make the process of launching an analysis easier. Infrequently used options are put on the options page and some phraseology has been improved. There had previously been some confusion with the class comparison tool about how to relate the output gene list to the three possible criteria for selecting genes (univariate p value, number of false discoveries, proportion of false discoveries). We changed the tool so that the user selects single criteria for each run.

Changes and Bug Fixes in Version 3.3.0:
==========

1:GenePix Importer: Added an option for background adjustment. Fixed the bug for reverse fluor data. Now, supports newer GenePix format.
2:Gene Subset Error: Fixed the bug in Gene subest option using CGAP and Biocarta and KEGG pathways.
3:Normalization: Fixed the bug for in print-tip Lowess normalization and housekeeping genes normalization for single channel
4:Class Prediction: Progress bar now works to indicate the time to run cross validation when permutation test is selected. Corrected the expression data table in the HTML output to get rid of an extraneous column.
5:PAM: Added a warning message about the impute function when more than 80% of the data is missing for an array.
6:GO Download: The utility has been removed due to extremely long download times and the latest release contains the most recent downloaded files.

7:Plugins: The following plugins may have passed incorrect data due to an Excel built-in function. 1-color data: Histogram and Smoothed CDF and 2 color data:ANOVA on log intensities, Histograms, Pairwise Correlation Plot, MA plot and Smoothed CDF.

What's New in Version 3.3

1) Enhanced heat map
    - more color coding options including multi-color rainbow
    - zoom in and out
    - labeling of genes

2) Pathway annotation of gene lists

3) Class comparison based on pathways rather than individual genes

4) Fast Fortran implementation of SAM
    - Approximately 7x faster than other implementations

5) Normalization of data separately by grid (print tip) for printed arrays

6) Direct import of GenePix data

7) Enhancements to Class Prediction analysis
    - Optimization of significance threshold for gene selection
    - New algorithm for selecting effective pairs of genes
    - Addition of shrunken centroid (PAM) classifier

8) New re-sampling methods for estimating prediction error
    - K-fold repeated cross-validation and .632 bootstrap options

9) Utility to compare gene lists

10) Plug-in for Random Forest classification

11) Plug-in for regression analysis of time series data to find regulated and differentially regulated genes

Changes and Bug Fixes in Version 3.2.3:

1) New plugin for regression analysis of time series data.

2) Fixed Fortran runtime error when running class comparison or class prediction using random variance model with more than 100 arrays. Previously, the run aborted without producing any output.

3) Fixed VBA runtime error in class comparison, class prediction, and various
analysis tools when user has more than 60,000 genes in the complete dataset.
Previously, the run aborted without producing any output.

4) Fixed error in class comparison where an empty string in the class label
was counted as a separate class in the analysis.

5) Fixed VBA runtime error in the utility to find intersection of genelists.

6) Fixed VBA runtime error in collation dialog for Affymetrix data archives
downloaded from the National Cancer Institutes's mAdb website.

7) Fixed VBA runtime error that occurred if user tried to click on the Plugins
menu item without an active workbook open in the Excel window.

Changes and Bug Fixes in Version 3.2.2:

1) Fixed the following bug in Class Comparison and Class Prediction tools
when the random variance option is selected and Affymetrix data is used:

```
Error in try(arr.current <- arr1[Filter==1,pheno==ClassLevels[1], :
 (subscript) logical subscript too long
Error occurred while executing the following R command:
 arr.current <- arr1[Filter==1,pheno==ClassLevels[1],drop=F]
```

Problem typically occurred whenever the yellow Filter column was one row
too long in the 'Filtered log intensity' and 'Gene identifiers' worksheets.
Problem was caused by a bug in Excel's built-in function to detect the last
used cell in a worksheet. A workaround has now been implemented to ensure
that the correct last cell can always be detected.

2) Fixed bug in Class Comparison Between Red and Green Channels, when some
of the experiments have been designated as reverse fluor arrays. Previously,
the Class Comparison results did not properly take the reverse fluor arrays
into account and flip the log-ratios back to their original values when
matching against the red and green class labels. This bug existed from the

3.2 Beta5 version onwards through the 3.2.1 versions.

3) Users who did not have the latest version of R installed encountered an
error message when trying to download the required Bioconductor libraries
for running the CEL file collation using RMA. The error message has now
been modified to direct users to download the latest R installer directly
from the CRAN website.

4) A minor modification was made in the Filter dialog on the 'Gene subsets'
page to retain the user's current selection of genelists even if the user
has added or deleted some genelist files from the file system.

Changes and Bug Fixes in Version 3.2.1:

1) Fixed bug in computation of number of unique permutations for randomized
block design in Class Comparison tool.

2) Added tool to find intersection between two genelists under Utilities menu.

3) No longer switches from old RServer.xla to new RExcel.xla, unless
ChangeRServer parameter is set to TRUE in the Preferences.txt file.
Switching from old RServer.xla to new RExcel.xla was causing mysterious
error messages for some users.

4) Gene subsetting parameters are now written to HTML output analysis files.

5) SOURCE annotations when using Gene Symbol as the lookup key now returns
more annotations. Previously, annotations were not returned whenever
the same Gene Symbol is represented in more than one organism (Human,
Mouse, or Rat).

6) Updated Affymetrix CEL file probe-level collation for compatibility with
R2.0.0.

7) Fixed cluster analysis dendrograms for compatibility with R2.0.0.

8) Added support for new Affymetrix chip types.

9) Fixed bug in downloading Gene Ontology structure when no project workbook
is opened.

10) Fixed bug in annotating HTML output analysis files when annotation contains
double-prime character or unmatched double-quotes.

Bug Fixes in Version 3.2 BETA_7:

1) Switching from old RServer.xla to new RExcel.xla caused some problems
passing arrays in Class Comparison and Class Prediction tools.

2) Switching from old RServer.xla to new RExcel.xla also caused analyses
to refilter project to ArrayTools default parameters, if user launched
Excel by directly opening a project workbook from the Windows Explorer,
rather than by opening Excel first BEFORE opening the project workbook.

3) FILTER variable was not properly passed to plugins.

Bug Fixes in Version 3.2 BETA_6:

1) Negative intensity value was set to missing rather than thresholded
for dual-channel data.

2) Sign of thresholded spots in reverse fluor arrays was not reversed.

3) Filtering and normalization was removed when gene subsets were selected
from the 'Select gene subsets' button at bottom of analysis dialogs for
single-channel data.

4) Housekeeping normalization of single-channel data incorrect when number
of arrays exceeded 15.

5) Fixed incorrect gene labels in Cluster Listing worksheet.

6) Fixed lowess smoother in M vs A (Log-ratio vs Avg log intensities)
scatterplot when array has large number of missing values.

Bug Fixes and Changes in Version 3.2 BETA_5:

1) Fixed error in list of genes selected by SAM tool.

2) Fixed random variance model when analysis is done on a gene subset. Previously, the inverse gamma parameters were based on gene subset rather than complete set of data, so that model assumptions were often not satisfied.

3) Fixed bug in running cluster reproducibility measures.

4) Added a menu item to subscribe or post to new ListServ

5) Changed labels on color legend in image plots (heat maps) to log scale.

Bug Fixes in Version 3.2 BETA_4:

1) Fixed SAM tool and added option to use 90th percentile instead of median for estimating confidence level for false discovery rate.

2) Fixed analyses related to Gene Ontology for Affymetrix data. Bug was caused by a format change in version 3.2 where the GO column is no longer written to the 'Gene ontology' worksheet for Affymetrix annotations.

3) Fixed Unigene and gene symbol hyperlinks in HTML output.

Bug Fixes in Version 3.2 BETA_3:

1) Fixed problems encountered in some large datasets when passing data to R in the analysis tools.

2) Fixed anonymous ftp for Windows NT users.

3) Fixed utility to download 'affy' package from BioConductor.

4) Changed multidimensional scaling to use maximum of 10 colors instead of 7.

5) Fixed automatic creation of Experiment Descriptors file so that array names can be read from horizontally aligned data.

What's New in Version 3.2

1) Automatic importation of Affymetrix CEL files. Calculation of Affymetrix
probe set summaries and normalization using RMA function of Bioconductor.

2) Importation of either log-transformed or not log-transformed data.

3) Class comparison to determine significance of Gene Ontology categories.

4) Class comparison to determine significance of user-defined genelists.

5) Extended class comparison for use in red-to-green comparisons with common reference.

6) Significance Analysis of Microarrays (SAM).

7) Speeded-up binary tree prediction tool using K-fold cross-validation.

8) ANOVA plugins tools:
- Fixed effect model for log-ratio or log-signal with up to 4 factors.
- Random effects model for log ratio or log signal.
- ANOVA for single channel intensities for dual-label arrays using non-common-reference design.

9) Optional parameters in 'Preferences.txt' file (in Prefs folder of ArrayTools installation folder) to control size of dendrogram plots produced by clustering tools.

10) Sample statistical considerations sections for publications included in Help documents.

11) Various bug fixes.

Feedback

Please send comments and bug reports to:

BRB-ArrayTools Development Team <arraytools@emmes.com>