Package 'ABarray'

December 23, 2024

Title Microarray QA and statistical data analysis for Applied Biosystems Genome Survey Microrarray (AB1700) gene expression data.

Version 1.74.0

Date 2006-02-11

Author Yongming Andrew Sun

Maintainer Yongming Andrew Sun <sunya@appliedbiosystems.com>

Imports Biobase, graphics, grDevices, methods, multtest, stats, tcltk, utils

Suggests limma, LPE

Description Automated pipline to perform gene expression analysis for Applied Biosystems Genome Survey Microarray (AB1700) data format. Functions include data preprocessing, filtering, control probe analysis, statistical analysis in one single function. A GUI interface is also provided. The raw data, processed data, graphics output and statistical results are organized into folders according to the analysis settings used.

biocViews Microarray, OneChannel, Preprocessing

License GPL

git_url https://git.bioconductor.org/packages/ABarray

git_branch RELEASE_3_20

git_last_commit ddd0046

git_last_commit_date 2024-10-29

Repository Bioconductor 3.20

Date/Publication 2024-12-23

Contents

`ABarray` *Utility to perform QA, data transformation and statistical analysis*

Description

(1) Read output from AB1700 software output; (2) Create raw data QA and associated plots including boxplot, control data signal plot; (3) Missing value calculation; (4) Create MA, scatter plot; (5) Perform quantile normalization; (6) Perform t test and fold change, or ANOVA (using separate function if more than 2 subgroups). (7) Create heatmap with hierarchical clustering. (8) The results are either in graphics or text files.

Usage

```
ABarray(dataFile, designFile, group, test = TRUE, impute = "avg", normMethod = "quantile", ...)
```

Arguments

`dataFile`	csv or tab delimit file contain expression measurement that are output from AB1700 software
`designFile`	Experiment design file, including information for sample type and additional phenotype information.
`group`	Specify which group statistical test will be performed on. The samples will be ordered according the group.
`test`	Specify whether to perform t test. By default, t test will be performed using specified group information.
`impute`	Treat flagged value (above 5000) as missing value, and impute the missing value.

`normMethod` The method of normalizaiton. The default is "quantile". The following normMethods are supported: quantile, mean, median, trimMean, and trimAMean. If the parameter value is one of the supported normMethods, the analysis will be performed on the chosen method. If the parameter value is "all", the analysis will be performed on quantile only, but the normalization results will be produced for each of the normMethods.

`...` Additional arguments. Use snThresh and/or detectSample to perform filtering. snThresh is the threshold of S/N value to be considered that the probe is detected (default value = 3, if snThresh is not specified). detectSample is used to determine if a probe should be included in statistical analysis (default value = 0.5, ie 50% of samples in any one subgroup).

Details

The function works on AB1700 software export data file. It expects certain file format to work. The rows of the file represent probes. The columns should contain these headings: probeID, geneID, Signal, S/N, Flag, and optionally SDEV, CV, AssayNormSignal (these values will be ignored in the process).

It is optional to have control probes. If they are present, plots will be generated for the control probes and they will be removed for further analysis.

It is required to have an experiment design file in certain format. The rows of the file are samples or arrays. The first column should be sampleName. Perhaps, sampleName should be concise and no spaces between characters. Second and third columns maybe assayName and arrayName (arrayName is optional). Additional columns should specify what type of samples. Note: It is best to have assayName the same as in dataFile.

Group name should be the same as in designFile. The samples will be ordered according the group information. The samples within the same subgroup will be ordered together. Only one group is accepted.

If test is TRUE (default), t test and ANOVA (if applicable) results will be produced.

If impute is avg (default), the signal values of the flagged probes will be imputed from average of the subgroup only if there are 2 or more values remaining in the subgroup.

Even if snThresh is not specified in the argument, snThresh is set to 3 by default. If a value other than 3 is desired (e.g., 2), put 'snThresh = 2' in the argument.

detectSample is also preset to a value = 0.5. This means that if a probe is detected in 50% or more samples in any subgroup within the group, it is included in statistical analysis. For example, if the group is named 'tissue', and there are 2 subgroups named 'lung' and 'liver', then, if a probe is detected in 50% or more samples in 'lung', it is included in the statistical analysis regardless the detectability in the other subgroup ('liver').

Value

An `ExpressionSet` object. The `assayDataElement(eset, "exprs")` will be populated with normalized signals, `assayDataElement(eset, "snDetect")` will be populated with S/N ratio values, and the `phenoData` slot will be populated with information from `designFile`. Further analysis can be performed on the `ExpressionSet` object with various R and Bioconductor packages.

Author(s)

Y Andrew Sun <sunya@appliedbiosystems.com>

See Also

doPlotEset, doPlotFCT, doANOVA, matrixPlot, mvaPair2, doLPE, doVennDiagram, hclusterPlot

Examples

```
#- eset <- ABarray(dataFile, designFile, "sampleGroup")
#- eset <- ABarray(dataFile, designFile, "group", detectSample = 0.8)
```

`ABarrayGUI` *GUI for ABarray to perform QA, data transformation and statistical analysis*

Description

A front end GUI for ABarray package to perform data analysis.

Usage

```
ABarrayGUI()
```

Details

The interface gathers required paramters for the ABarray packages to run. See ABarray for more details.

Value

No return values.

Author(s)

Y Andrew Sun <sunya@appliedbiosystems.com>

See Also

ABarray, doPlotEset, doPlotFCT, doANOVA, matrixPlot, mvaPair2, doLPE, doVennDiagram, hclusterPlot

Examples

```
#- ABarrayGUI()
```

`calcsn` *Calculate SN summary for each group*

Description

Calculate S/N ratio summary for each group

Usage

```
calcsn(sn, snThresh, pdata, group, grpMember)
```

Arguments

`sn`	S/N ratio data
`snThresh`	S/N threshold filtering
`pdata`	experiment design
`group`	which group should be calculated
`grpMember`	optional, members of the group

Value

data matrix

Author(s)

Y Andrew Sun

`concord` *Calculate signal detection concordance*

Description

Calculate signal detection concordance between columns using S/N threshold (default = 3)

Usage

```
concord(sn, snThresh = 3)
```

Arguments

`sn`	a matrix containing s/n ratio
`snThresh`	S/N threshold to use, default = 3

Value

a matrix with the concordance

Author(s)

Y Andrew Sun

Examples

```
#-concordance <- concord(sn)  ##- sn ratio matrix
```

`cvv` *CV calculation*

Description

Calculate cv

Usage

```
cvv(data)
```

Arguments

`data`	data matrix contain expression values

Value

vector of cv for each gene or probe

Author(s)

Yongming Sun

`cvvPlot` *Plot CV value*

Description

Plot CV value against average intensity (log2)

Usage

```
cvvPlot(data, name)
```

Arguments

`data`	vector of cv for each gene
`name`	name of the plot

Value

None

Author(s)

Yongming Sun

See Also

cvv cvv to calulate cv

doANOVA *Perform one way or two way ANOVA*

Description

If only one factor is provided in parameter, one way ANOVA is performed. If two factors are provided, two way ANOVA is performed.

Usage

```
doANOVA(eset, group1, group2, snThresh = 3, detectSample = 0.5)
```

Arguments

eset	An ExpressionSet object.
group1	A factor name or labels to test on. If eset is an ExpressionSet object, either name or labels can be used. If eset is an expression matrix, labels should be used.
group2	A factor name or labels to test on.
snThresh	Using probes detectable for ANOVA analysis, default S/N value is 3 or more to be considered detectable.
detectSample	The percentage of samples the probe is detected in order to be considered in ANOVA analysis.

Details

At least one group should be provided. If ExpressionSet object is used, group1 or group2 is the name of the sampleGroup defined in experiment design file. If labels are to be used, they can be either numeric or text, e.g., c(1,1,2,2,3,3) or c("treat1", "treat1", "treat2", "treat2", "treat3", "treat3").

If the probe is detectable in 50% (default) or more samples in any one of the subgroup, it is included in the ANOVA analysis.

Value

a vector if one way ANOVA; a matrix if two way ANOVA

Author(s)

Y Andrew Sun

Examples

```
  #- one way ANOVA
#-  anova <- doANOVA(eset, "sampleGroup")

  #- two way ANOVA
#-  anova <- doANOVA(eset, "sampleGroup1", "sampleGoup2")
```

doLPE *Perform LPE analysis*

Description

The local pooled error test attempts to reduce dependence on the within-gene estimates in tests for differential expression, by pooling error estimates within regions of similar intensity. Note that with the large number of genes there will be genes with low within-gene error estimates by chance, so that some signal-to-noise ratios will be large regardless of mean expression intensities and fold-change. The local pooled error attempts to avert this by combining within-gene error estimates with those of genes with similar expression intensity.

Usage

```
doLPE(eset, group, member, name = "", snThresh = 3, detectSample = 0.5)
```

Arguments

eset	an ExpressionSet object
group	which group should LPE be performed
member	optional. The member names in the group specified above
name	a prefix name for use when writing output to file
snThresh	S/N ratio threshold to use to define gene detectability
detectSample	percentage of samples detectable above snThresh to include in LPE test. The default is 50%. If the probe is detected in 50% or more samples in one of the subgroup, it is considered in LPE analysis

Details

The LPE test statistic numerator is the difference in medians between the two experimental conditions. The test statistic denominator is the combined pooled standard error for the two experimental conditions obtained by looking up the var.M from each baseOlig.error variance function. The conversion to p-values is based on the Gaussian distribution for difference if order statistics (medians). The user may select both the smoother degrees of freedom (smaller is smoother) and the trim percent to obtain a variance function to suit particular issues i.e. variability of genes with low expression intensity.

Value

Dataframe

Author(s)

Y Andrew Sun

References

Bioconductor LPE package

Examples

```
##---- Some example usage ----
```

doPlotEset *Produce a number of QA plot plus t and ANOVA test*

Description

Produce boxplot, MA plot, scatter plot, correlation, S/N detection concordance, CV, and t test, ANOVA test if subgroup is more than 2

Usage

```
doPlotEset(eset, group, name = "", snThresh = 3, test = TRUE, ...)
```

Arguments

eset	an ExpressionSet object
group	name of the group from experiment design file
name	a name for use in output files for record purpose
snThresh	threshold of S/N considered detectable, default = 3
test	whether t or ANOVA test should be performed
...	Additional arguments, currently not implemented

Details

The t test and fold change is performed with function fctPlot. See additional information with fctPlot. ANOVA is performed with doANOVA.

If there are more than 2 subgroup in group, t test and fold change will be performed for each pair of subgroup and one way ANOVA will be performed. If subgroup is 2, ANOVA will not be performed.

Value

None. A number of plots and t or ANOVA test result file will be produced.

Author(s)

Y Andrew Sun

Examples

```
#-doPlotEset(eset, "sampleGroup")
#-doPlotEset(eset, "sampleGroup", name = "perfect")
#-doPlotEset(eset, "sampleGroup", test = FALSE)  ##- t test will be not performed
```

doPlotFCT *Calculate fold change and t test, the plot*

Description

Calculate fold changes and p values from t test, and plot the results using preset FDR threshold

Usage

```
doPlotFCT(eset, group, grpMember, order1 = NULL, order2 = NULL,
detectSample = 0.5, snThresh = 3, ...)
```

Arguments

eset	an ExpressionSet object
group	which group from experiment design should calculation and plot be performed
grpMember	optional group member within the group
order1	optional, For a pairwise comparison the ordering of the first group of replicates
order2	optional, For a pairwise comparison the ordering of the first group of replicates
detectSample	optional number between 0 and 1 to indicate the percentage of arrays should be above snThresh to include in the t test analysis. Default = 0.5. If the probe is detected in 50% or more samples on one of the subgroup, the probe is included in the t test, otherwise, it will be excluded in the t test
snThresh	optional S/N ratio threshold. Default = 3
...	Additional argument, currently not implemented

Details

Group members are optional. For example, if group name is "tissue", and group members in experiment design file include "brain", "liver", "lung", "muscle". We could include c("brain", "liver") as group member for the parameter, then t test will be performed between "brain" and "liver", and "lung" "muscle" will be ignored. However, if we omit group member in the arguments, all tissue members will be used for t test. In this case, there will be 6 pairwise t test between each member of the group.

If order1 and order2 are specified then a paired sample t-test will be conducted between the groups, with the arrays in each group sorted according to the ordering specified. For example, if order1 is c(1,3,2) and order2 is c(1,2,3), then the sample pairing is a1-b1, a3-b2, a2-b3, with a and b are subgroup 1 and subgroup 2 within the group.

The fold changes are difference between averaged subgroup1 expression vs averaged subgroup2. If paired t test is performed, the fold changes are calculated using each paired difference and take an average of paired difference.

Value

None. But a number of plot and result files will be produced.

Author(s)

Y Andrew Sun

Examples

```
#- doPlotFCT(eset, "sampleGroup", c("liver", "muscle"))
#- For a paired t test
#- doPlotFCT(eset, "sampleGroup", c("liver", "muscle"), order1 = c(1,2,3), order2 = c(1,3,2))
```

`doVennDiagram` *Create Venn Diagram*

Description

Create Venn diagram from lists.

Usage

```
doVennDiagram(a, b, c = NULL, names, ...)
```

Arguments

a	a vector of first list
b	a vector of second list
c	a vector of third list, optional
names	a vector for the name of the set
...	additional graphical parameter

Details

The funciton will create Venn diagram. If two lists (a and b) are provided, two-way Venn diagram will produced. If three lists (a, b, and c) are provided, three-way Venn diagram will be produced.

This function depends on some functions of limma package, and is derived from limma package.

Value

A plot of Venn diagram

Author(s)

Yongming Sun

References

Bioconductor limma package.

drawVennDiagram *Draw Venn Diagram*

Description

Drawing actual Venn diagram

Usage

```
drawVennDiagram(object, names, mar = rep(0.5, 4), cex = 1, ...)
```

Arguments

object	VennCounts object produced by VennCounts, which is numeric matrix with last column "Counts" giving counts for each possible vector outcome
names	optional character vector giving names for the sets
mar	numeric vector of length 4 specifying the width of the margins around the plot. This argument is passed to par.
cex	numerical value giving the amount by which the contrast names should be scaled on the plot relative to the default.plotting text. See par.
...	any other arguments are passed to plot

Value

a plot of Venn Diagram

Author(s)

Yongming Sun

References

Bioconductor Limma package

Examples

```
##---- Do not call this function directly !! ----
```

getMemberEset *Produce a sub ExpressionSet given a group and its members*

Description

From a group and its member name, return an ExpressionSet containing just these members

Usage

```
getMemberEset(eset, group, member)
```

Arguments

eset	an ExpressionSet object
group	the name of the group which must be in the experiment design file
member	member name(s) in the above mentioned group

Value

an ExpressionSet object

Author(s)

Yongming Sun

getPantherMap *Create pie chart for probes involved in Panther Pathway*

Description

Given a list of probeID, attempt to find out panther classification information

Usage

```
getPantherMap(probeID, title, figDir)
```

Arguments

probeID	a list of probeIDs
title	the title for the figure to be generated
figDir	directory for the figures to be placed in

Value

None. Several figures will be generated.

Author(s)

Yongming Sun

hclusterPlot *heatmap generation*

Description

plot clustering heatmap using correlation

Usage

```
hclusterPlot(expr, title, dist)
```

Arguments

expr	matrix of gene expression value
title	the title for the plot
dist	whether to use correlation or distance for clustering, default to use Euclidean distance. Use dist = "Correlation" to cluster with correlation coefficient

Details

generating heatmap using correlation as distance

Value

None. heatmap will be generated.

Author(s)

Y Andrew Sun

icpPlot *icp plot function*

Description

QC plot for internal control probes

Usage

```
icpPlot(controlData, colProbeID = 1, plotWhat = "Signal", pdfDir, jpgDir)
```

Arguments

controlData	Signal intensity matrix for icp probes
colProbeID	the column where probeID is located
plotWhat	Whether we are plotting signal or S/N
pdfDir	a directory where pdf files should be produced
jpgDir	a directory where jpg or bmp files should be produced

Value

A series of QC plots

Author(s)

Yongming Sun

Examples

```
##---- Do not call this function DIRECTLY !! ----
```

---

imputeFlag *Perform imputation for missing values (FLAG > 5000)*

Description

Perform imputation for missing values.

Usage

```
imputeFlag(rawSig, pd = NULL, group = "", impute = "avg")
```

Arguments

rawSig	a matrix containing gene expression with missing values labeled as NA
pd	phenoData object
group	which group should average be performed
impute	choice of impute method, only avg (average) is implemented

Value

a list containing a matrix with the imputed values and rows that are imputed.

Author(s)

Y Andrew Sun

Examples

```
#-imputed <- imputeFlag(raw, pd, group = "tissue", impute = "avg")  ##- sn ratio matrix
```

lpe.fdr.BH *Perform FDR on LPE results*

Description

Perform Benjamini and Hochberg FDR adjustment on LPE results

Usage

```
lpe.fdr.BH(lpe.result, adjp = "BH")
```

Arguments

lpe.result	the result from LPE analysis
adjp	Type of adjustment, default "BH"

Details

Do not call this function directly. Called from doLPE

Value

a matrix with original and ajusted p values

Author(s)

Yongming Sun

References

Bioconductor LPE package

Examples

```
##---- Do not call this function directly !! ----
```

mamaplot *MA plot function*

Description

plot MA from vectors A and M

Usage

```
mamaplot(A, M, idx, subset = sample(1:length(M), min(c(10000, length(M)))), span = 2/3, family.loess
```

Arguments

A	vector of average signal
M	vector of difference signal
idx	index for which S/N < 3
subset	subset
span	span
family.loess	loess fit
cex	cex value
...	additional arguments

Value

MA plot

Note

Modified from bioconductor affy package

Author(s)

Yongming Sun

References

bioconductor affy package

See Also

See Also as mvaPair2

Examples

```
##---- Do not call this function DIRECTLY !! ----
```

---

matrixPlot *heatmap for matrix*

Description

Create heatmap from a matrix

Usage

```
matrixPlot(x, nrgcols = 50, rlabels = TRUE, clabels = TRUE, rcols = 1, ccols = 1, k = 10, title = "", .
```

Arguments

x	a matrix
nrgcols	number of colors to use
rlabels	whether to use row labels
clabels	whether to use column labels
rcols	use supplemental row label
ccols	use supplemental column label
k	number of tick labels for scale bar
title	title for the plot
...	additional argument

Details

This function can be used to plot any numberic matrix, e.g., correlation matrix, S/N matrix, signal intensity matrix, etc

Value

heatmap

Author(s)

Yongming Sun

---

mvaPair2 *plot MA for each pair of columns*

Description

MA plot for each pair of columns

Usage

```
mvaPair2(x, y = NULL, snThresh = 3, labels = colnames(x), log.it = FALSE, span = 2/3,
    family.loess = "gaussian", digits = 3, line.col = 2, main = "MA plot", ... )
```

Arguments

x	expression matrix
y	S/N ratio matrix
snThresh	S/N threshold
labels	name for the labels
log.it	should data be log transformed
span	span of the plot
family.loess	curve fitting
digits	number of digits to display
line.col	size of the line col
main	title for the MA plot
...	additional argument

Details

If S/N ratio is available, probes with S/N < 3 in both array will be colored differently.

Value

MA plot

Author(s)

Yongming Sun

Examples

```
##---- exprs expression matrix, sn s/n ratio !! ----
```

---

panel.cor *Create correlation panel*

Description

Create correlation panel

Usage

```
panel.cor(x, y, digits=3, prefix="", cex.cor)
```

Arguments

x	vector of expression value for one sample
y	vector of expression value for another sample
digits	number of digits to display the correlation
prefix	additional text to display
cex.cor	size of the text

Value

None

Author(s)

Yongming Sun

Examples

```
##---- Not intended for direct function call !! ----
```

`panel.scatter` *Creat scatter plot*

Description

Create scatter plot

Usage

```
panel.scatter(x, y, col = "blue", bg = NA, pch = ".",
    cex = 1, col.smooth = "red", span = 2/3, iter = 3, ...)
```

Arguments

`x`	vector of expression for one sample
`y`	vector of expression for another sample
`col`	color of points
`bg`	background colors
`pch`	pch paremeter
`cex`	size of text
`col.smooth`	color of smooth line
`span`	span of the plot
`iter`	iteration
`...`	additional arguments

Value

None

Author(s)

Yongming Sun

Examples

```
##---- Not intended for use this function directly !! ----
```

qnNormalize *Perform quantile normalization*

Description

Perform quantile normalization between arrays

Usage

```
qnNormalize(eData, snr, method = 'quantile', snThresh = 3, ties = TRUE)
```

Arguments

eData	matrix of gene expression values
snr	Optional signal/noise ratio. Only used for trimAMean method
method	The normalization method desired. Default method is quantile
snThresh	Signal/noise threshold (default = 3) to indicate presence or absence of a probe signal
ties	handle values with same rank

Details

This function performs various normalization for the array data. The default is quantile normalization method (adapted from Bioconductor limma package). Other normalization methods include median, mean, trimMean (trimmed mean), trimAMean (mean with absent gene removed).

For the median normalizaiton, the median signal of each array is scaled to the same value (this value is calculated to equal to the median of all values in the data). The signal values for each array are then adjusted by the scaling factor.

For the mean normalization, the approach is similar to the median normalization procedure except that the mean signal of each array is scaled to the same value (this value is median of all signals in the data).

For the trimMean normalization, the approach is similar to the mean normalization except that the mean for each array is calculated after trimming the top and botton 5% of signals (a total of 10% of values).

For the trimAMean normalization, the signal values for absent probes are not considered. If the s/n of a probe is less than snThresh (default = 3), the expression of the probe is considered not present (absent). The remaining values are then trimmed (top and botton 2.5%, a total of 5%), and the mean value for each array after trimming is scaled to the same value (median of all values in the data).

Value

data matrix with quantile normalized data values

Author(s)

Yongming Sun

References

bioconductor limma package for quantile normalization

rgcolorsfunc *generate color map*

Description

Generate color map for heatmap use

Usage

```
rgcolorsfunc(n = 50)
```

Arguments

n	number of colors to generate

Value

rgb color vector

Author(s)

Yongming Sun

Examples

```
## Do not call this function directly
rgb <- rgcolorsfunc()
```

savejpg *save device to jpg image file*

Description

save plot device to jpg image file

Usage

```
savejpg(x, width = 1024, height = 768)
```

Arguments

x	file name to be saved to
width	The width for the figure in pixal
height	The height for the figure

Value

For windows version, it produce bmp formatted image, otherwise, produce jpg images.

Author(s)

Yongming Sun

scaleColorBar *Create scale for heatmap*

Description

Create a bar for heatmap scales

Usage

```
scaleColorBar(x, horizontal = FALSE, col = rgcolorsfunc(50), scale = 1:length(x),
    k = 10, cLen = 9, ...)
```

Arguments

x	vector of scales need to be plotted
horizontal	whether the bar is vertical or horizontal
col	color function
scale	scale of the bar
k	number of intervals on scale
cLen	length of columns
...	additional arguments

Value

none

Author(s)

Yongming Sun

Examples

```
##--- Do not call this function directly !! ----
```

snSummary *Create summary information for S/N ratio*

Description

Create summary information for S/N ratio for each sample group

Usage

```
snSummary(eset, snThresh = 3, group, grpMember)
```

Arguments

eset	an ExpressionSet object
snThresh	S/N ratio threshold to use, default = 3
group	sample group
grpMember	sample group members, optional

Value

a matrix containing the number of samples with S/N >=3 for each probe

Author(s)

Yongming Sun

Index