Package 'NASHMAP'

April 15, 2016

Title Microarray Analysis Pipeline

Version 1.0

Description NASHMAP is an R package to normalize, analyze, and visualize microarray data

Depends R (>= 3.2.0)

Imports limma,
affy,
samr,
tools,
stringr,
DESeq2,
gplots,
ggplot2,
RColorBrewer

License GPL

LazyData true

RoxygenNote 5.0.1

R topics documented:

`backgroundCutoff` *Visualize arrays to identify cutoff for background subtraction*

Description

Plot the histogram of max gene illumination across all arrays to identify cutoff for background subtraction. Upper and lower limits can be changed to narrow focus. Plots will be saved in analysis pipeline directory.

Usage

```
backgroundCutoff(preprocessData.obj, method=c("mloess","quantile"),
xlim.lo=0, xlim.hi)
```

Arguments

preprocessData.obj	Object returned from call to preprocessData
method	Type of normalization to use: "quantile" for quantile normalization; "mloess" for MLOESS normalization
xlim.lo	Lower bound on X for histogram plot (binary logarithm)
xlim.hi	Upper bound on X for histogram plot (binary logarithm)

---

backgroundSubtraction *Perform background subtraction on normalized data*

Description

Remove all genes that have an maximum intensity level below that of the provided cutoff.

Usage

```
backgroundSubtraction(preprocessData.obj,
method=c("mloess","quantile"), cutoff)
```

Arguments

preprocessData.obj	Object returned from call to preprocessData
method	Type of normalization to use: "quantile" for quantile normalization; "mloess" for MLOESS normalization
cutoff	The cutoff value (binary logarithm) identified through backgroundCutoff

Value

A list with components

ntext	Number of leading text columns
dataCol	Vector of column indices containing array data
id	Vector containing gene ID's
idInd	Column index containing gene ID information
symbol	Vector containing gene symbols
symbolIndex	Column index containing gene symbol information
descStats	Vector of column indices containing descriptive statistics
pipelineName	Name of pipeline generated from input file name sans extension
data	Data frame of descriptive stats and normalized data for chosen method

pathwayHeatmap *Create heatmaps on desired subsets of genes*

Description

Normalize input raw data using quantile and mloess methods. Plots of the normalized data along with a dendrogram clustering all samples will be stored in newly created pipeline directory.

Usage

```
preprocessPlots(runSAM.obj, pathwaysDir, fileFormat=c("geneid",
"symbol"))
```

Arguments

runSAM.obj	Object returned from call to runSAM
pathwaysDir	Directory containing files of genes output from pathway analysis
fileFormat	Indicator of how genes are identified in each file, be it "geneid" or "symbol"

preprocessData *Generate MA plots of raw and normalized data*

Description

Normalize input raw data using quantile and mloess methods. Plots of the normalized data along with a dendrogram librarclustering all samples will be stored in newly created pipeline directory.

Usage

```
preprocessPlots(inputFile, fileSheet=1, ntext=2, dataCol, symbolIndex=1,
 idIndex=2)
```

Arguments

inputFile	Path to the microarray expression file, be it .xlsx or .csv
fileSheet	Sheet number in the spreadsheet with data
ntext	Number of leading text columns
dataCol	Range of columns which contain data (indexing begins with first column of file)
symbolIndex	Column index which contains gene symbols
idIndex	Column index which contains gene ID's

Value

A list with components

ntext	Number of leading text columns
dataCol	Vector of column indices containing array data
id	Vector containing gene ID's
idIndex	Column index containing gene ID information
symbol	Vector containing gene symbols
symbolIndex	Column index containing gene symbol information
descStats	Vector of column indices containing descriptive statistics
pipelineName	Name of pipeline generated from input file name sans extension
mloess	Data rame of quantile normalized data
quantile	Data rame of quantile normalized data

---

runSAM *Identify significant genes through SAM*

Description

Implement SAM and compute significant genes given delta. Output will consist of all significant genes ordered by increasing q-value and decreasing d-score.

Usage

```
runSAM(backgroundSub.obj, classCompareCols, classCompareName,
fdr.cutoff=0.1, response)
```

Arguments

backgroundSub.obj	Object returned from call to backgroundSub
classCompareCols	Vector of column indices indicating which subset of arrays are to be compared for this comparison
classCompareName	String title given to the name of the comparison
fdr.cutoff	Max FDR for SAM, will use delta value which results in max FDR below this cutoff
response	Vector of 1, 2 values that indicate group membership

Value

A list with components

`siggenesTable`	Combined data frame of genes having significant positive and negative correlation
`dataCol`	Vector of column indices containing array data
`ntext`	Number of leading text columns
`response`	Vector of array group membership, 1=control, 2=experimental
`pipelineName`	Name of pipeline generated from input file name sans extension
`data`	Data frame of chosen normalization method data
`classCompareCols`	Value entered through classCompareCols parameter
`classCompareName`	Value entered through classCompareName parameter
`symbolIndex`	Column index that contains gene symbol

Index