Package 'diffGeneAnalysis'

November 11, 2026

Title Performs differential gene expression Analysis

Version 1.94.0

Author Choudary Jagarlamudi

Description Analyze microarray data

Maintainer Choudary Jagarlamudi <choudary.jagar@swosu.edu>

License GPL

Collate dataTrim.R curveFit.R normalize.R biasAdjust.R assocAnalysis.R refGroup.R

Imports graphics, grDevices, minpack.lm (>= 1.0-4), stats, utils

biocViews Microarray, DifferentialExpression

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

git_branch RELEASE_3_23

git_last_commit 083d069

git_last_commit_date 2026-04-28

Repository Bioconductor 3.23

Date/Publication 2026-11-11

Contents

assocAnalysis 2
biasAdjust 3
curveFit 4
dataTrim 5
normalize 6
rawdata 7
refGroup 8

Index **11**

assocAnalysis *Associative Analysis*

Description

assocAnalysis performs the Associative-T method of differential gene analysis.

Usage

```
assocAnalysis(bAdjusted, numSlides, ctrl, exp, sdrefgrp,
pctrl, pexpm,ctrlavjs, expmavjs, ctrlsds, expmsds, rgrplen)
```

Arguments

bAdjusted	bAdjusted is a matrix of bias adjusted normalized microarray data.
numSlides	numSlides represents the total number of chips in the microarray experiment.
ctrl	ctrl is the number of control chips in the microarray experiment.
exp	exp is the number of control chips in the microarray experiment.
sdrefgrp	standard deviation of the computed refernce group.
pctrl	p values of the control chips on performin ttests.
pexpm	p values of the experiment chips on performin ttests.
ctrlavjs	averages of genes across all control chips.
expmavjs	averages of genes across all experiment chips.
ctrlsds	standard deviations of genes across all control chips.
expmsds	standard deviations of genes across all experiment chips.
rgrplen	total number of genes of the computed refernce group.

Details

AssociativeAnalysis performs the Associative-T method of differential gene analysis.The results are displayed in a 10 column matrix as follows

Column Representation.

1 Gene Bank Id

2 Average Signal of the Control Chips/Channels.

3 SD of Control Chips/Channels.

4 Probability that a given gene in the Control Chips/Channels belongs or does not belong to background.

5 Average Signal of the Experimental Chips/Channels.

6 SD of Experimental Chips/Channels.

7 Probability that a given gene in the Experimental Chips/Channels belongs or does not belong to background.

8 P-value from a Student T-test.

9 P-value from an Associative T-test.

10 Ratio of mean expression values (Control/Experimental).

11 Group Number.

Group Numbers are defined as follows:.

A1 Expressed above background in both sample types, but over-expressed on the Experimental Chips/Channels.

A2 Expressed above background in both sample types, but over-expressed on the Control Chips/Channels.

A3 Expressed above background only on the Experimental Chips/Channel.

A4 Expressed above background only in the Control Chips/Channel.

0 None of the above.

Value

A matrix of 11 columns as described in the details and n rows where n stands for the number of genes in the microarray dataset.

Author(s)

Choudary L Jagarlamudi

References

Dozmorov I,Centola,M. An associative analysis of gene expression array data.Bioinformatics.2003 Jan22;19(2):204-11

Knowlton N,Dozmorov I, Centola M. Microarray data Analysis Tool box(MDAT): for normalization,adjustment and analysis of gene expression data.Bioonformatics. 2004 Dec 12;20(18):3687-90

Examples

```
#see refGroup for more details
```

---

biasAdjust *baisAdjust performs bias adjustment of normalized data*

Description

biasAdjust takes a normalized dataset and applies a multiplicative scalar derived from the data to help account for expression biases. These expression biases can come from many sources including dye bias,hybridization efficacy,changes in personnel,etc.After bias adjustment the data is ready for differential gene analysis.

Usage

```
biasAdjust(normalized, numSlides)
```

Arguments

normalized	normalized is a matrix of normalized microarray data.The first row consists of headers ans the first column consists of gene names.
numSlides	numSlides is the total number of chips of the microarray dataset.

Value

Returns a matrix of bias adjusted normalized data.

Note

biasAdjust takes 2-3 seconds to execute under optimal conditions of size of datasets and speed of the machine..Tested on a Pentium IV 1.6Ghz, 256Mb RAM with 22464 genes and 10 chips it took 12 seconds to execute.

Author(s)

Choudary L Jagarlamudi

References

Dozmorov I,Centola,M. An associative analysis of gene expression array data. Bioinformatics.2003 Jan22;19(2):204-11

Knowlton N,Dozmorov I, Centola M. Microarray data Analysis Tool box(MDAT): for normalization,adjustment and analysis of gene expression data. Bioonformatics.2004 Dec 12;20(18):3687-90

Examples

```
#biasAdjust(normalized,7)
```

---

curveFit *CurveFit data to a Gaussian distribution*

Description

CurveFit takes a vector of chipdata from microarray slides and fits the data to a Gaussian distribution through a non-linear least-squares optimization algorithm.The results are graphically depicted in a series of histograms. Each histogram represents a different initial seed (left to right: 2 bins, 3 bins, 4 bins, 4.5 bins, 5 bins, and 5.5 bins) that is passed to the curve fitting algorithm. The resulting fit for each histogram is superimposed with a solid blue line.The user is then able to visually select the 'best' fit.

Usage

```
curveFit(chipdata, plot)
```

Arguments

chipdata	a vector of chipdata from microarray slides.
plot	plot can take values of 1 or 0. If plot is 1 then the histogram with the curve fit will be shown graphically.

Value

an object res which is a list containing the following components. res1]: mean of the computed background. res[2]: standard deviation of the computed background.

Author(s)

Choudary L Jagarlamudi

References

Dozmorov I,Centola,M. An associative analysis of gene expression array data. Bioinformatics.2003 Jan22;19(2):204-11

Knowlton N,Dozmorov I, Centola M. Microarray data Analysis Tool box(MDAT): for normalization,adjustment and analysis of gene expression data. Bioonformatics.2004 Dec 12;20(18):3687-90

Examples

```
#see normalize for details.
```

---

dataTrim *DataTrim using pseudo winsorization algorithm*

Description

DataTrim cleans the data through a pseudo winsorization algorithm. First, the mean and SD are calculated. Then, any values above 2SD are trimmed.The mean is recalculated and any values less then 2 SD are trimmed.This process of cutting 2 SD above the mean and then 2 SD below the mean is repeated until no further cuts are possible.

Usage

```
dataTrim(chipdata)
```

Arguments

chipdata	chipdata is a vector of chipdata from microarray chips.

Value

Returns a vector of trimmed chipdata representing background based on pseudo winsorization algorithm.

Author(s)

Choudary L Jagarlamudi

References

Dozmorov I,Centola,M. An associative analysis of gene expression array data. Bioinformatics.2003 Jan22;19(2):204-11

Knowlton N,Dozmorov I, Centola M. Microarray data Analysis Tool box(MDAT): for normalization,adjustment and analysis of gene expression data. Bioonformatics.2004 Dec 12;20(18):3687-90

Examples

```
#see normalize for details
```

`normalize` *Normalization of microarray data*

Description

Normalization of data utilizing information obtained from background fluoresence.Background fluoresce intensity values are used to determine a Gaussian distribution of lowly expressed genes,yielding the background estimates(mean and standard deviation).

Usage

```
normalize(rawdata, numSlides,ctrl,expm,ctrlbg,expmbg)
```

Arguments

`rawdata`	rawdata is matrix of microarray data.The first column consists of gene Names and the first row consists of headers.
`numSlides`	numSlides represents the total number of chips/slides in the microarray dataset including control and experiment.Control slides are always followed by experiment slides from left to right in the matrix.
`ctrl`	ctrl represents the total number of control chips in the microarray dataset.
`expm`	expm represents the total number of experiment chips in the microarray dataset.
`ctrlbg`	ctrlbg represents the percent of data to pick for background computation of the control chips.30 percent is the default.
`expmbg`	expmbg represents the percent of data to pick for background computation of the experiment chips.30 percent is the default.

Details

The normalization algorithm trims the data based on innitial emperical estimates of the mean and standard deviation.All data beyond +/-2SD of the mean are cut iteratively.This procedure is repeated until no more cuts can be made.The trimmed data is then subjected to a non linear curve fitting procedure. The user is presented with six different pictures obtained using bars 2,3,4,4.5,5,and5.5 as mean. The user is given the freedom to select the best visual estimate of background. The user selected parameters are used to perform a z-Trnasformation on the data.The percent of data selected to compute background depends on the data obtained.The default is 30 percent.A normal distributed histogram should confirm that, else the user is allowed to pick a percent and make changes until the user sees a normal distributed histogram.Upon running normalize the user is presented with a set of 6 histograms. If the user is not happy with the default 30 percent, the user should go ahead and select a mean and confirm curvefit,then select 'no' to confirm histogram distribution.The user will be presented with a new set of 6 histograms. This process is repeated until the user selects the best Histogram distribution.This process is repeated for each individual chip.

Value

A matrix of normalized values of rawdata

Author(s)

Choudary L Jagarlamudi

References

Dozmorov I,Centola,M. An associative analysis of gene expression array data. Bioinformatics.2003 Jan22;19(2):204-11

Knowlton N,Dozmorov I, Centola M. Microarray data Analysis Tool box(MDAT): for normalization,adjustment and analysis of gene expression data. Bioonformatics.2004 Dec 12;20(18):3687-90

Examples

```
#rawdata is loaded in the package. Run example as follows:
#Read the description file for best results.
#data(rawdata)
#normalize(rawdata,7,3,4,0.15,0.60)
```

`rawdata` *Micro array dataset*

Description

Microarray dataset consisting of 7 chips.3 control and 4 experiment with 2382 genes.The first column consists of gene names, hence the dataset has 8 columns in total.

Usage

```
data(rawdata)
```

Format

A data frame with 2382 observations on the following 11 variables.

GeneID Gene IDS or Gene names

c1 a numeric vector of gene intensities for control chip 1

c2 a numeric vector of gene intensities for control chip 2

c3 a numeric vectorof gene intensities for control chip 3

e1 a numeric vector of gene intensities for experiment chip 1

e2 a numeric vector of gene intensities for experiment chip 2

e3 a numeric vector of gene intensities for experiment chip 3

e4 a numeric vector of gene intensities for experiment chip 4

Source

Dozmorov I,Centola,M. An associative analysis of gene expression array data. Bioinformatics.2003 Jan22;19(2):204-11

Knowlton N,Dozmorov I, Centola M. Microarray data Analysis Tool box(MDAT): for normalization,adjustment and analysis of gene expression data. Bioonformatics.2004 Dec 12;20(18):3687-90

Examples

```
data(rawdata)
```

---

refGroup *Reference Group computes a robust estimate of inter-assay variability (Standard Error)*

Description

refGroup takes a normalized, bias adjusted dataset and computes a robust estimate of inter-assay variability (Standard Error).This value is used here to perform T-tests. It can also be used in sample size calculations.The associative analysis method is then applied to the computed refernce group.

Usage

```
refGroup(biasAdjusted, numSlides, ctrl, exp, pval)
```

Arguments

`biasAdjusted`	biasAdjusted is a matrix of bias adjusted normalized microarray data.The first row consists of headers and the first column must consist of Gene Bank Id's.
`numSlides`	numSlides represents the total number of chips in a microarray dataset including control and experiment.
`ctrl`	ctrl represents the total number of ctrl chips in the microarray.
`exp`	exp represents the total number of experiment chips in the microarray.
`pval`	pval is the strigency value for computing the reference group. 0.05 is the normally suggested value.

Details

AssociativeAnalysis performs the Associative-T method of differential gene analysis.The user is asked to enter values for E and R.E stand for the increase in fold over background and R stands for the ratio of experimental chips average over control chips average. The higher these values the higher will be the stringency. Example dataset used here was run with an E values of 1 and R value of 1.5.

The results are displayed in a 10 column matrix as follows

Column Representation.

1 Gene Bank ID

2 Average Signal of the Control Chips/Channels.

3 SD of Control Chips/Channels.

4 Probability that a given gene in the Control Chips/Channels belongs or does not belong to background.

5 Average Signal of the Experimental Chips/Channels.

6 SD of Experimental Chips/Channels.

7 Probability that a given gene in the Experimental Chips/Channels belongs or does not belong to background.

8 P-value from a Student T-test.

9 P-value from an Associative T-test.

10 Ratio of mean expression values (Control/Experimental).

11 Group Number.

Group Numbers are defined as follows:.

A1 Expressed above background in both sample types, but over-expressed on the Experimental Chips/Channels.

A2 Expressed above background in both sample types, but over-expressed on the Control Chips/Channels.

A3 Expressed above background only on the Experimental Chips/Channel.

A4 Expressed above background only in the Control Chips/Channel.

0 None of the above.

Value

A matrix of 11 columns as described in the details and n rows where n stands for the number of genes in the microarray dataset.

Note

refGroup takes 4-5 seconds to execute under optimal conditions of size of datasets and speed of the machine.Tested on Pentium IV 1.6Ghz, 256 MB RAM with 22464 genes and 10 chips it took 25 seconds to execute.

Author(s)

Choudary L Jagarlamudi

References

Dozmorov I,Centola,M. An associative analysis of gene expression array data.Bioinformatics. 2003 Jan22;19(2):204-11

Knowlton N,Dozmorov I, Centola M. Microarray data Analysis Tool box(MDAT): for normalization,adjustment and analysis of gene expression data.Bioonformatics. 2004 Dec 12;20(18):3687-90

Examples

```
#refGroup(bAdjusted,7,3,4,0.05)
```

Index

∗ **internal**
assocAnalysis, 2
biasAdjust, 3
curveFit, 4
dataTrim, 5
normalize, 6
rawdata, 7
refGroup, 8

assocAnalysis, 2

biasAdjust, 3

curveFit, 4

dataTrim, 5

normalize, 6

rawdata, 7
refGroup, 8