Package 'DRI'

February 14, 2012

Type Package

Title DR-Integrator

Version 1.1

Date 2009-11-16

Author Keyan Salari, Robert Tibshirani, Jonathan R. Pollack

Maintainer Keyan Salari <ksalari@stanford.edu>

Description Integrative analysis of DNA copy number and gene expression data described in Salari et al 2009

License GPL (>= 2)

Depends impute, cghFLasso

LazyLoad yes

Repository CRAN

Date/Publication 2009-11-17 08:55:54

R topics documented:

DRI-package 2
drcorrelate 3
drcorrelate.null 5
dri.correlation.plot 6
dri.fdrCutoff 7
dri.heatmap 9
dri.impute 10
dri.merge.CNbyRNA 12
dri.sig_genes 13
dri.smooth.cghdata 14
drsam 15
drsam.null 17
mySampleData 18
runFusedLasso 19

Index 21

DRI-package *DR-Integrator: an analytic tool for integrating DNA copy number and gene expression data*

Description

DR-Integrator identifies genes with significant correlations between DNA copy number alterations and gene expression data, and implements a supervised learning analysis that captures genes with significant alterations in both DNA copy number and gene expression between two sample classes.

Details

Package:	DRI
Type:	Package
Version:	1.1
Date:	2009-11-16
License:	GPL-2

This package contains two analytic tools: DR-Correlate and DR-SAM.

Author(s)

Keyan Salari, Robert Tibshirani, Jonathan R. Pollack

Maintainer: Keyan Salari <ksalari@stanford.edu>

References

Salari, K., Tibshirani, R., and Pollack, J.R. (2009) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. http://pollacklab.stanford.edu/

See Also

drcorrelate, drcorrelate.null, drsam, drsam.null, dri.fdrCutoff, dri.sig_genes, dri.heatmap, dri.merge.CNbyRNA, dri.smooth.cghdata, runFusedLasso

Examples

```
require(impute)
data(mySampleData)
attach(mySampleData)

# DNA data should contain no missing values - pre-smooth beforehand
# Impute missing values for gene expression data
RNA.data <- dri.impute(RNA.data)
```

```
# DR-Correlate analysis to find genes with correlated DNA/RNA measurements
obs <- drcorrelate(DNA.data, RNA.data, method="pearson")
# generate null distribution for FDR calculation (10 permutations)
null <- drcorrelate.null(DNA.data, RNA.data, method="pearson", perm=10)
# identify the correlation cutoff corresponding to your desired FDR
n.cutoff <- dri.fdrCutoff(obs, null, targetFDR=0.05, bt=TRUE)
cutoff <- n.cutoff[2]
# retrieve all genes that are significant at the determined cutoff, and
# calculate gene-specific FDRs
Results <- dri.sig_genes(cutoff, obs, null, GeneIDs, GeneNames, Chr, Nuc,
bt=TRUE, method="drcorrelate")

# Optional heatmap plot for significant DR-Correlation genes
sample.names <- colnames(DNA.data)
pdf(file="DRI-Heatmap.pdf", height=8, width=11)
dri.heatmap(Results, DNA.data, RNA.data, sample.names, GeneNames, Chr, Nuc,
statistic="pearson", color.scheme="RG")
dev.off()

# DR-SAM analysis to find genes with alterations in both DNA and RNA between
# different classes
labels <- c(rep(1,25), rep(2,25)) # 25 samples in class 1 and 25 in class 2
obs <- drsam(DNA.data, RNA.data, labels, transform.type="raw")
# generate null distribution for FDR calculation (10 permutations)
null <- drsam.null(DNA.data, RNA.data, labels, transform.type="raw", 10)
# identify the correlation cutoff corresponding to your desired FDR
n.cutoff <- dri.fdrCutoff(obs$test.summed, null, targetFDR=0.05, bt=TRUE)
cutoff <- n.cutoff[2]
# retrieve all genes that are significant at the determined cutoff, and
# calculate gene-specific FDRs
Results <- dri.sig_genes(cutoff, obs, null, GeneIDs, GeneNames, Chr, Nuc,
bt=TRUE, method="drsam")
```

`drcorrelate` *Compute correlations between DNA copy number and gene expression for each gene*

Description

A correlation is computed between each gene's DNA copy number and gene expression across all the samples. Significant correlations are determined by comparison to a null distribution derived from random permutations of the data.

Usage

```
drcorrelate(DNA, RNA, method = "pearson", tail_p = 10)
```

Arguments

DNA	matrix of DNA copy number data
RNA	matrix of gene expression data, samples (columns) in same order as DNA matrix
method	correlation statistic, either "pearson", "spearman", or "ttest"
tail_p	top/bottom percent of samples (with respect to the gene's copy number) to use for extremes t-test groups; used only when method = "ttest"

Details

DR-Correlate aims to identify genes with expression changes explained by underlying CNAs. This tool performs an analysis to identify genes with statistically significant correlations between their DNA copy number and gene expression levels. Three options for the statistic to measure correlation are implemented: (1) Pearson's correlation; (2) Spearman's rank correlation; and (3) an "extremes" t-test. For Pearson's and Spearman's correlations, the respective correlation coefficient is computed for each gene. For the extremes t-test, a modified Student's t-test (Tusher et al. 2001) is computed for each gene, comparing gene expression levels of samples comprising the lowest and highest deciles with respect to DNA copy number. That is, for each gene the samples are rank ordered by DNA copy number and samples below the 10th percentile and above the 90th percentile form two groups whose means of gene expression are compared with a modified t-test. The percentile cutoff defining the two groups is user-adjustable.

Value

observed	a vector of observed correlations for each gene

Author(s)

Keyan Salari, Robert Tibshirani, and Jonathan R. Pollack

References

Salari, K., Tibshirani, R., and Pollack, J.R. (2009) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. http://pollacklab.stanford.edu/

See Also

drcorrelate, drcorrelate.null, drsam, drsam.null, dri.fdrCutoff, dri.sig_genes, dri.heatmap, dri.merge.CNbyRNA, dri.smooth.cghdata, runFusedLasso

Examples

```
require(impute)
data(mySampleData)
attach(mySampleData)

# DNA data should contain no missing values - pre-smooth beforehand
# Impute missing values for gene expression data
RNA.data <- dri.impute(RNA.data)
```

```
# DR-Correlate analysis to find genes with correlated DNA/RNA measurements
obs <- drcorrelate(DNA.data, RNA.data, method="pearson")
# generate null distribution for FDR calculation (10 permutations)
null <- drcorrelate.null(DNA.data, RNA.data, method="pearson", perm=10)
# identify the correlation cutoff corresponding to your desired FDR
n.cutoff <- dri.fdrCutoff(obs, null, targetFDR=0.05, bt=TRUE)
cutoff <- n.cutoff[2]
# retrieve all genes that are significant at the determined cutoff, and
# calculate gene-specific FDRs
Results <- dri.sig_genes(cutoff, obs, null, GeneIDs, GeneNames, Chr, Nuc,
bt=TRUE, method="drcorrelate")
```

---

drcorrelate.null *Generate null distribution for DR-Correlate analysis*

Description

A null distribution is generated by randomly permuting the sample labels of the gene expression data matrix and recomputing the DNA/RNA correlations for each gene.

Usage

```
drcorrelate.null(DNA, RNA, method = "pearson", tail_p = 10, perm)
```

Arguments

DNA	matrix of DNA copy number data
RNA	matrix of gene expression data, samples (columns) in same order as DNA matrix
method	correlation statistic, either "pearson", "spearman", or "ttest"
tail_p	top/bottom percent of samples (with respect to the gene's copy number) to use for extremes t-test groups; used only when method = "ttest"
perm	number of permutations to perform

Value

null	n * k matrix of null data, where n = number of genes and k = number of permutations

Author(s)

Keyan Salari, Robert Tibshirani, and Jonathan R. Pollack

References

Salari, K., Tibshirani, R., and Pollack, J.R. (2009) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. http://pollacklab.stanford.edu/

See Also

drcorrelate, drcorrelate.null, drsam, drsam.null, dri.fdrCutoff, dri.sig_genes, dri.heatmap, dri.merge.CNbyRNA, dri.smooth.cghdata, runFusedLasso

Examples

```
require(impute)
data(mySampleData)
attach(mySampleData)

# DNA data should contain no missing values - pre-smooth beforehand
# Impute missing values for gene expression data
RNA.data <- dri.impute(RNA.data)

# DR-Correlate analysis to find genes with correlated DNA/RNA measurements
obs <- drcorrelate(DNA.data, RNA.data, method="pearson")
# generate null distribution for FDR calculation (10 permutations)
null <- drcorrelate.null(DNA.data, RNA.data, method="pearson", perm=10)
# identify the correlation cutoff corresponding to your desired FDR
n.cutoff <- dri.fdrCutoff(obs, null, targetFDR=0.05, bt=TRUE)
cutoff <- n.cutoff[2]
# retrieve all genes that are significant at the determined cutoff, and
# calculate gene-specific FDRs
Results <- dri.sig_genes(cutoff, obs, null, GeneIDs, GeneNames, Chr, Nuc,
bt=TRUE, method="drcorrelate")
```

---

dri.correlation.plot *Plot significant correlations from DR-Correlate analysis*

Description

A plot is generated of all the correlations computed by DR-Correlate, indicating significant correlations in red (positive) or blue (negative), ordered by chromosomal coordinates.

Usage

```
dri.correlation.plot(observed, Results.SigGenes, sig_cutoff, chr,
nuc_pos, bothtails)
```

Arguments

observed	vector of observed correlations from drcorrelate
Results.SigGenes	list of significant genes from dri.sig_genes
sig_cutoff	correlation significance cutoff returned from dri.fdrCutoff
chr	vector of gene chromosome locations
nuc_pos	vector of gene nucleotide positions
bothtails	TRUE or FALSE indicating whether 2-tail test was performed

Value

plot — a plot of the correlations is returned in chromosomal order, with significant correlations marked in red (positive) and blue (negative)

Author(s)

Keyan Salari, Robert Tibshirani, Jonathan R. Pollack

References

Salari, K., Tibshirani, R., and Pollack, J.R. (2009) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. http://pollacklab.stanford.edu/

See Also

drcorrelate, drcorrelate.null, drsam, drsam.null, dri.fdrCutoff, dri.sig_genes, dri.heatmap, dri.merge.CNbyRNA, dri.smooth.cghdata, runFusedLasso

Examples

```
require(impute)
data(mySampleData)
attach(mySampleData)

# DNA data should contain no missing values - pre-smooth beforehand
# Impute missing values for gene expression data
RNA.data <- dri.impute(RNA.data)

# DR-Correlate analysis to find genes with correlated DNA/RNA measurements
obs <- drcorrelate(DNA.data, RNA.data, method="pearson")
# generate null distribution for FDR calculation (10 permutations)
null <- drcorrelate.null(DNA.data, RNA.data, method="pearson", perm=10)
# identify the correlation cutoff corresponding to your desired FDR
n.cutoff <- dri.fdrCutoff(obs, null, targetFDR=0.05, bt=TRUE)
cutoff <- n.cutoff[2]
# retrieve all genes that are significant at the determined cutoff, and
# calculate gene-specific FDRs
Results <- dri.sig_genes(cutoff, obs, null, GeneIDs, GeneNames, Chr, Nuc,
bt=TRUE, method="drcorrelate")

# Optional correlation plot for significant DR-Correlation genes
dri.correlation.plot(obs, Results, cutoff, Chr, Nuc, bothtails=TRUE)
```

---

dri.fdrCutoff — *Determine cutoff score for a desired false discovery rate (FDR)*

Description

A search is performed for the cutoff score (for either DR-Correlate or DR-SAM) that corresponds to the user-defined FDR.

Usage

```
dri.fdrCutoff(observed, null, targetFDR, bt = TRUE)
```

Arguments

`observed`	vector of scores from either `drcorrelate` or `drsam`
`null`	matrix of null data from either `drcorrelate.null` or `drsam.null`
`targetFDR`	desired false discovery rate
`bt`	either `TRUE` or `FALSE` indicating whether a 2-tail test was performed

Details

A binary search is implemented to find the cutoff score that corresponds to the user-defined FDR

Value

`comp1`	a two element list containing the number of genes found significant at the chosen FDR, and the score cutoff corresponding to that FDR.

Author(s)

Keyan Salari, Robert Tibshirani, and Jonathan R. Pollack

References

Salari, K., Tibshirani, R., and Pollack, J.R. (2009) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. http://pollacklab.stanford.edu/

See Also

`drcorrelate`, `drcorrelate.null`, `drsam`, `drsam.null`, `dri.fdrCutoff`, `dri.sig_genes`, `dri.heatmap`, `dri.merge.CNbyRNA`, `dri.smooth.cghdata`, `runFusedLasso`

Examples

```
require(impute)
data(mySampleData)
attach(mySampleData)

# DNA data should contain no missing values - pre-smooth beforehand
# Impute missing values for gene expression data
RNA.data <- dri.impute(RNA.data)

# DR-Correlate analysis to find genes with correlated DNA/RNA measurements
obs <- drcorrelate(DNA.data, RNA.data, method="pearson")
# generate null distribution for FDR calculation (10 permutations)
null <- drcorrelate.null(DNA.data, RNA.data, method="pearson", perm=10)
# identify the correlation cutoff corresponding to your desired FDR
n.cutoff <- dri.fdrCutoff(obs, null, targetFDR=0.05, bt=TRUE)
cutoff <- n.cutoff[2]
```

```
# retrieve all genes that are significant at the determined cutoff, and
# calculate gene-specific FDRs
Results <- dri.sig_genes(cutoff, obs, null, GeneIDs, GeneNames, Chr, Nuc,
bt=TRUE, method="drcorrelate")

# DR-SAM analysis to find genes with alterations in both DNA and RNA between
# different classes
labels <- c(rep(1,25), rep(2,25)) # 25 samples in class 1 and 25 in class 2
obs <- drsam(DNA.data, RNA.data, labels, transform.type="raw")
# generate null distribution for FDR calculation (10 permutations)
null <- drsam.null(DNA.data, RNA.data, labels, transform.type="raw", 10)
# identify the correlation cutoff corresponding to your desired FDR
n.cutoff <- dri.fdrCutoff(obs$test.summed, null, targetFDR=0.05, bt=TRUE)
cutoff <- n.cutoff[2]
# retrieve all genes that are significant at the determined cutoff, and
# calculate gene-specific FDRs
Results <- dri.sig_genes(cutoff, obs, null, GeneIDs, GeneNames, Chr, Nuc,
bt=TRUE, method="drsam")
```

---

dri.heatmap *Generate a heatmap of significant DR-Correlate genes*

Description

A heatmap is generated showing the DNA copy number and gene expression data for the significant, positively correlated DR-Correlate genes.

Usage

```
dri.heatmap(Results.SigGenes, DNA, RNA, SampleIDs, GeneNames, Chr,
Nuc, statistic, color.scheme)
```

Arguments

Results.SigGenes	list of significant genes from `dri.sig_genes`
DNA	matrix of DNA copy number data
RNA	matrix of gene expression data, samples (columns) in same order as DNA matrix
SampleIDs	vector of sample names
GeneNames	vector of gene names
Chr	vector of gene chromosome locations
Nuc	vector of gene nucleotide positions
statistic	method used in `drcorrelate`, either "pearson", "spearman", or "ttest"
color.scheme	desired heatmap color scheme, either "RG" (red-green), "RB" ("red-blue), or "YB" (yellow-blue)

Author(s)

Keyan Salari, Robert Tibshirani, Jonathan R. Pollack

References

Salari, K., Tibshirani, R., and Pollack, J.R. (2009) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. http://pollacklab.stanford.edu/

See Also

drcorrelate, drcorrelate.null, drsam, drsam.null, dri.fdrCutoff, dri.sig_genes, dri.heatmap, dri.merge.CNbyRNA, dri.smooth.cghdata, runFusedLasso

Examples

```
require(impute)
data(mySampleData)
attach(mySampleData)

# DNA data should contain no missing values - pre-smooth beforehand
# Impute missing values for gene expression data
RNA.data <- dri.impute(RNA.data)

# DR-Correlate analysis to find genes with correlated DNA/RNA measurements
obs <- drcorrelate(DNA.data, RNA.data, method="pearson")
# generate null distribution for FDR calculation (10 permutations)
null <- drcorrelate.null(DNA.data, RNA.data, method="pearson", perm=10)
# identify the correlation cutoff corresponding to your desired FDR
n.cutoff <- dri.fdrCutoff(obs, null, targetFDR=0.05, bt=TRUE)
cutoff <- n.cutoff[2]
# retrieve all genes that are significant at the determined cutoff, and
# calculate gene-specific FDRs
Results <- dri.sig_genes(cutoff, obs, null, GeneIDs, GeneNames, Chr, Nuc,
bt=TRUE, method="drcorrelate")

# Optional heatmap plot for significant DR-Correlation genes
sample.names <- colnames(DNA.data)
pdf(file="DRI-Heatmap.pdf", height=8, width=11)
dri.heatmap(Results, DNA.data, RNA.data, sample.names, GeneNames, Chr, Nuc,
statistic="pearson", color.scheme="RG")
dev.off()
```

---

dri.impute *Impute missing values for gene expression data*

Description

Missing gene expression measurements are imputed by the K-nearest neighbors method

Usage

```
dri.impute(d)
```

Arguments

d	data matrix whose missing values will be imputed

Details

The K-nearest neighbors method is employed (K=10) to impute missing gene expression values. For any row with more than 50

Uses the `impute.knn` function of the `impute` package.

Value

data	the new imputed data matrix

Author(s)

Keyan Salari, Robert Tibshirani, and Jonathan R. Pollack

References

Salari, K., Tibshirani, R., and Pollack, J.R. (2009) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. http://pollacklab.stanford.edu/

Olga Troyanskaya, Michael Cantor, Gavin Sherlock, Pat Brown, Trevor Hastie, Robert Tibshirani, David Botstein and Russ B. Altman, Missing value estimation methods for DNA microarrays BIOINFORMATICS Vol. 17 no. 6, 2001 Pages 520-525.

See Also

`drcorrelate`, `drcorrelate.null`, `drsam`, `drsam.null`, `dri.fdrCutoff`, `dri.sig_genes`, `dri.heatmap`, `dri.merge.CNbyRNA`, `dri.smooth.cghdata`, `runFusedLasso`

Examples

```
require(impute)
data(mySampleData)
attach(mySampleData)

# Impute missing values for gene expression data
RNA.data <- dri.impute(RNA.data)
```

dri.merge.CNbyRNA *Merge DNA copy number and gene expression data sets*

Description

A matrix of DNA copy number values and a matrix of gene expression values are merged into one matrix using the genome position coordinates of the probes on each microarray platform. Copy number probes flanking each gene expression probe are averaged to derive the copy number at a given expression probe.

Usage

```
dri.merge.CNbyRNA(dna.chr, dna.nuc, dna.data, rna.chr, rna.nuc)
```

Arguments

`dna.chr`	vector of chromosomes on which copy number probes are located
`dna.nuc`	numeric vector of nucleotide positions where copy number probes are located
`dna.data`	matrix of DNA copy number data
`rna.chr`	vector of chromosomes on which expression probes are located
`rna.nuc`	numeric vector of nucleotide positions where expression probes are located

Details

This function performs a pre-processing step to obtain 1-to-1 mappings of DNA copy number and gene expression for each measured gene. For each gene expression probe, the closest 5' and 3' copy number probes are used to calculate an average copy number value. If the same array platform was used to measure copy number and gene expression, the copy number and gene expression measurement will be matched by the probes' genome position coordinates. This function returns a DNA copy number matrix that has the same number of rows (genes) as the gene expression matrix.

Value

`average.dna.data`
a matrix of average DNA copy number values corresponding to each gene expression value

Author(s)

Keyan Salari, Robert Tibshirani, and Jonathan R. Pollack

References

Salari, K., Tibshirani, R., and Pollack, J.R. (2009) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. http://pollacklab.stanford.edu/

See Also

drcorrelate, drcorrelate.null, drsam, drsam.null, dri.fdrCutoff, dri.sig_genes, dri.heatmap, dri.merge.CNbyRNA, dri.smooth.cghdata, runFusedLasso

Examples

```
require(impute)
data(mySampleData)
attach(mySampleData)

dri.merge.CNbyRNA(dna.chr=Chr, dna.nuc=Nuc, dna.data=DNA.data, rna.chr=Chr,
rna.nuc=Nuc)
```

---

dri.sig_genes *Determine significant genes for DR-Correlate or DR-SAM analysis*

Description

Given a cutoff score corresponding to the desired FDR, the list of significant genes is generated, along with each gene's FDR.

Usage

```
dri.sig_genes(cutoff, observed, null_dist, gene_id, gene_name, chr,
nuc, bt = TRUE, method = "drcorrelate")
```

Arguments

cutoff	cutoff score for significance from dri.fdrCutoff
observed	vector of observed scores from drcorrelate or drsam
null_dist	matrix of null data from drcorrelate.null or drsam.null
gene_id	vector of gene IDs
gene_name	vector of gene names
chr	vector of gene chromosome locations
nuc	vector of gene nucleotide positions
bt	either TRUE or FALSE indicated whether a 2-tail test was performed
method	analysis method used, either "drcorrelate" or "drsam"

Value

Results.SigGenes
a list of significant genes, positive and negative, with gene-specific FDRs

Author(s)

Keyan Salari, Robert Tibshirani, and Jonathan R. Pollack

References

Salari, K., Tibshirani, R., and Pollack, J.R. (2009) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. http://pollacklab.stanford.edu/

See Also

drcorrelate, drcorrelate.null, drsam, drsam.null, dri.fdrCutoff, dri.sig_genes, dri.heatmap, dri.merge.CNbyRNA, dri.smooth.cghdata, runFusedLasso

Examples

```
require(impute)
data(mySampleData)
attach(mySampleData)

# DNA data should contain no missing values - pre-smooth beforehand
# Impute missing values for gene expression data
RNA.data <- dri.impute(RNA.data)

# DR-Correlate analysis to find genes with correlated DNA/RNA measurements
obs <- drcorrelate(DNA.data, RNA.data, method="pearson")
# generate null distribution for FDR calculation (10 permutations)
null <- drcorrelate.null(DNA.data, RNA.data, method="pearson", perm=10)
# identify the correlation cutoff corresponding to your desired FDR
n.cutoff <- dri.fdrCutoff(obs, null, targetFDR=0.05, bt=TRUE)
cutoff <- n.cutoff[2]
# retrieve all genes that are significant at the determined cutoff, and
# calculate gene-specific FDRs
Results <- dri.sig_genes(cutoff, obs, null, GeneIDs, GeneNames, Chr, Nuc,
bt=TRUE, method="drcorrelate")
```

---

dri.smooth.cghdata *Smooth DNA copy number data over a specific window size*

Description

A moving average is calculated by chromosomal location of copy number probes using a specified window size.

Usage

```
dri.smooth.cghdata(DNA.data, Chr, mw = 5)
```

Arguments

DNA.data	matrix of DNA copy number data (rows ordered by genome position
Chr	vector of chromosomes on which copy number probes are located (ordered by probes' genome position
mw	window size to take moving average over

Details

This function performs one pre-processing method for smoothing copy number data before performing further DNA/RNA integrative analyses. To help eliminate missing values, and reduce noise, for each sample a moving average is taken over a user-specified window size along each chromosome.

Value

`DNA.smoothed` matrix of DNA copy number data smoothed by a moving average window

Author(s)

Keyan Salari, Robert Tibshirani, and Jonathan R. Pollack

References

Salari, K., Tibshirani, R., and Pollack, J.R. (2009) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. http://pollacklab.stanford.edu/

See Also

`drcorrelate`, `drcorrelate.null`, `drsam`, `drsam.null`, `dri.fdrCutoff`, `dri.sig_genes`, `dri.heatmap`, `dri.merge.CNbyRNA`, `dri.smooth.cghdata`, `runFusedLasso`

Examples

```
data(mySampleData)
attach(mySampleData)

DRI.DNA.processed <- dri.smooth.cghdata(DNA.data, Chr, 5)
```

---

drsam   *Perform supervised learning analysis between 2 sample classes for DNA/RNA differences*

Description

A test is performed to identify genes with significant differences in both DNA copy number and gene expression between two sample groups of interest.

Usage

```
drsam(DNA.data, RNA.data, labels, transform.type, for.null = FALSE)
```

Arguments

DNA.data	matrix of DNA copy number data
RNA.data	matrix of gene expression data, samples (columns) in same order as DNA matrix
labels	class labels of the two comparison groups, either 1 or 2
transform.type	type of transformation to apply to data, either "standardize", "rank", or "raw"
for.null	used internally by drsam.null, keep as default

Details

DR-SAM (DNA/RNA-Significance Analysis of Microarrays) performs a supervised analysis to identify genes with statistically significant differences in both DNA copy number and gene expression between different classes (e.g., tumor subtype-A vs. tumor subtype-B). The goal of this analysis is to identify genetic differences (CNAs) that mediate gene expression differences between two groups of interest. DR-SAM implements a modified Student's t-test to generate for each gene two t-scores assessing differences in DNA copy number and differences in gene expression. A final score is computed by first summing the copy number t-score and gene expression t-score, and then weighting the sum by the ratio of the two t-scores. The weight is applied to favor genes with strong differences in both DNA copy number and gene expression between the two classes.

Value

observed	vector of observed DR-SAM scores for each gene

Author(s)

Keyan Salari, Robert Tibshirani, Jonathan R. Pollack

References

Salari, K., Tibshirani, R., and Pollack, J.R. (2009) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. http://pollacklab.stanford.edu/

See Also

drcorrelate, drcorrelate.null, drsam, drsam.null, dri.fdrCutoff, dri.sig_genes, dri.heatmap, dri.merge.CNbyRNA, dri.smooth.cghdata, runFusedLasso

Examples

```
require(impute)
data(mySampleData)
attach(mySampleData)

# DNA data should contain no missing values - pre-smooth beforehand
# Impute missing values for gene expression data
RNA.data <- dri.impute(RNA.data)

# DR-SAM analysis to find genes with alterations in both DNA and RNA between
# different classes
```

```
labels <- c(rep(1,25), rep(2,25)) # 25 samples in class 1 and 25 in class 2
obs <- drsam(DNA.data, RNA.data, labels, transform.type="raw")
# generate null distribution for FDR calculation (10 permutations)
null <- drsam.null(DNA.data, RNA.data, labels, transform.type="raw", 10)
# identify the correlation cutoff corresponding to your desired FDR
n.cutoff <- dri.fdrCutoff(obs$test.summed, null, targetFDR=0.05, bt=TRUE)
cutoff <- n.cutoff[2]
# retrieve all genes that are significant at the determined cutoff, and
# calculate gene-specific FDRs
Results <- dri.sig_genes(cutoff, obs, null, GeneIDs, GeneNames, Chr, Nuc,
bt=TRUE, method="drsam")
```

drsam.null — *Generate null distribution for DR-SAM analysis*

Description

A null distribution is generated by randomly permuting the class labels of the samples and recomputing the DR-SAM statistic for each gene.

Usage

```
drsam.null(DNA.data, RNA.data, labels, transform.type, k)
```

Arguments

`DNA.data`	matrix of DNA copy number data
`RNA.data`	matrix of gene expression data, samples (columns) in same order as DNA matrix
`labels`	class labels of the two comparison groups, either 1 or 2
`transform.type`	type of transformation to apply to data, either "standardize", "rank", or "raw"
`k`	number of permutations to perform

Value

`null`	n * k matrix of null data, where n = number of genes and k = number of permutations

Author(s)

Keyan Salari, Robert Tibshirani, Jonathan R. Pollack

References

Salari, K., Tibshirani, R., and Pollack, J.R. (2009) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. http://pollacklab.stanford.edu/

See Also

drcorrelate, drcorrelate.null, drsam, drsam.null, dri.fdrCutoff, dri.sig_genes, dri.heatmap, dri.merge.CNbyRNA, dri.smooth.cghdata, runFusedLasso

Examples

```
require(impute)
data(mySampleData)
attach(mySampleData)

# DNA data should contain no missing values - pre-smooth beforehand
# Impute missing values for gene expression data
RNA.data <- dri.impute(RNA.data)

# DR-SAM analysis to find genes with alterations in both DNA and RNA between
# different classes
labels <- c(rep(1,25), rep(2,25)) # 25 samples in class 1 and 25 in class 2
obs <- drsam(DNA.data, RNA.data, labels, transform.type="raw")
# generate null distribution for FDR calculation (10 permutations)
null <- drsam.null(DNA.data, RNA.data, labels, transform.type="raw", 10)
# identify the correlation cutoff corresponding to your desired FDR
n.cutoff <- dri.fdrCutoff(obs$test.summed, null, targetFDR=0.05, bt=TRUE)
cutoff <- n.cutoff[2]
# retrieve all genes that are significant at the determined cutoff, and
# calculate gene-specific FDRs
Results <- dri.sig_genes(cutoff, obs, null, GeneIDs, GeneNames, Chr, Nuc,
bt=TRUE, method="drsam")
```

---

mySampleData *DR-Integrator Sample Data*

Description

Sample data of 100 genes copy number and gene expression data from 50 cell lines

Usage

```
data(mySampleData)
```

Format

The format is: List of 6 $ GeneIDs : Factor $ GeneNames: Factor $ Chr : int $ Nuc : int $ DNA.data : matrix $ RNA.data : matrix

Source

Kao, J., Salari, K., Bocanegra, M.C. et al. (2009) Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. PLoS One.

References

Salari, K., Tibshirani, R., and Pollack, J.R. (2009) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. http://pollacklab.stanford.edu/

Examples

```
data(mySampleData)
```

---

runFusedLasso — *Wrapper function to run cghFLasso*

Description

A function to run the Fused Lasso method on DNA copy number data to call copy number alterations.

Usage

```
runFusedLasso(DNA.data, normal.data = NA, chr, nuc, FDR)
```

Arguments

`DNA.data`	matrix of disease tissue DNA copy number data
`normal.data`	matrix of normal tissue DNA copy number data (optional)
`chr`	vector of chromosomes where copy number probes are located (ordered by genome location)
`nuc`	numeric vector of nucleotide positions where copy number probes are located (ordered by genome location)
`FDR`	false discovery rate to apply to threshold significant gains and losses

Details

This function performs a pre-processing step to smooth and segment DNA copy number data using the fused lasso method of Tibshirani & Wang, 2008. The DNA copy number data from the disease tissues are compared to that of normal tissues, if supplied, to help call significant copy number alterations. A false discovery rate is calculated for each copy number alteration and the significance threshold is user-definable. More details of the Fused Lasso method can be found with the `cghFLasso` package and in the reference below.

Value

`data.FL`	matrix of DNA copy number data with significant gains and losses called by Fused Lasso, and non-significant changes set to zero

Author(s)

Keyan Salari, Robert Tibshirani, and Jonathan R. Pollack

References

Salari, K., Tibshirani, R., and Pollack, J.R. (2009) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. http://pollacklab.stanford.edu/

Tibshirani, R., and Wang, P. (2008) Spatial smoothing and hot spot detection for CGH data using the fused lasso, Biostatistics, 9, 18-29.

See Also

`drcorrelate`, `drcorrelate.null`, `drsam`, `drsam.null`, `dri.fdrCutoff`, `dri.sig_genes`, `dri.heatmap`, `dri.merge.CNbyRNA`, `dri.smooth.cghdata`, `runFusedLasso`

Examples

```
library(cghFLasso)
data(CGH)
attach(CGH)

DRI.DNA.data.FL <- runFusedLasso(DNA.data=DiseaseArray,
normal.data=NormalArray, chr=chromosome, nuc=nucposition, FDR=0.01)
```

Index

∗Topic **datasets**
mySampleData, 18
∗Topic **package**
DRI-package, 2

drcorrelate, *2*, 3, *4*, *6–8*, *10, 11*, *13–16*, *18*, *20*
drcorrelate.null, *2*, *4*, 5, *6–8*, *10, 11*, *13–16*, *18*, *20*
DRI *(*DRI-package*)*, 2
DRI-package, 2
dri.correlation.plot, 6
dri.fdrCutoff, *2*, *4*, *6*, 7, *7, 8*, *10, 11*, *13–16*, *18*, *20*
dri.heatmap, *2*, *4*, *6–8*, 9, *10, 11*, *13–16*, *18*, *20*
dri.impute, 10
dri.merge.CNbyRNA, *2*, *4*, *6–8*, *10, 11*, 12, *13–16*, *18*, *20*
dri.sig_genes, *2*, *4*, *6–8*, *10, 11*, 13, *13–16*, *18*, *20*
dri.smooth.cghdata, *2*, *4*, *6–8*, *10, 11*, *13*, 14, *14–16*, *18*, *20*
drsam, *2*, *4*, *6–8*, *10, 11*, *13, 14*, 15, *15, 16*, *18*, *20*
drsam.null, *2*, *4*, *6–8*, *10, 11*, *13–16*, 17, *18*, *20*

mySampleData, 18

runFusedLasso, *2*, *4*, *6–8*, *10, 11*, *13–16*, *18*, 19, *20*