Package 'GeneExpressionSignature'

October 7, 2014

Title Gene Expression Signature based Similarity Metric

Version 1.10.0

Date 2012-10-24

Author Yang Cao

Maintainer Yang Cao <yiluheihei@gmail.com>, Fei Li <pittacus@gmail.com>,Lu Han <hanl8910@gmail.com>

Description This package gives the implementations of the gene expression signature and its distance to each. Gene expression signature is represented as a list of genes whose expression is correlated with a biological state of interest. And its distance is defined using a nonparametric, rank-based pattern-matching strategy based on the Kolmogorov-Smirnov statistic. Gene expression signature and its distance can be used to detect similarities among the signatures of drugs,diseases, and biological states of interest.

Depends R (>= 2.13), Biobase, PGSEA

Suggests apcluster,GEOquery

License GPL-2

LazyLoad yes

biocViews GeneExpression

R topics documented:

`exampleSet` *sample data, a subset of the C-MAP*

Description

sample data, a subset of the C-MAP as , which is a collection of 50 genome-wide transcriptional expression data from cultured human cells treated with 15 different small molecules.

Usage

```
data(exampleSet)
```

Format

A ExpressionSet: assay data represents the 50 genome-wide transcriptional expression data, phenotypic data describes 15 different small molecules corresponds to the expression data (assay data).

References

http://www.sciencemag.org/content/313/5795/1929.short Lamb et al., The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease, science 2006

Examples

```
data(exampleSet)
```

`GeneExpressionSignature`

Introduction to the GeneExpressionSignature Package

Description

The **GeneExpressionSignature** add-on is an implementation of computing distances among preprocessed gene-expression profiles of samples for R. The distances can be used to detect similarities among the signatures of drugs, diseases, and biological states of interest, and construct connectivity map.

Details

This package contains functions for the distances computation based on gene expression signature. First, list of genes is ranked according to their expression ratios to produce the Prototype Ranked List (PRL). Second, all the PRLs with the same state are aggregated by `RankMerging` functions. Finally, all the ranked lists are made as one input of the `ScorePGSEA` and `ScoreGSEA` functions to compute the pairwise distances.

getRLs *Convert gene expression profiles to a ranked list (data preprocessing)*

Description

Sorting the microarray probe-set identifiers according to the differential expression values with respect to the untreated hybridization to obtaine a ranked list. Gene-expression profiles in are represented in a nonparametric fashion.

Usage

```
getRLs(control, treatment)
```

Arguments

control	a matrix, including the vehicle control gene expression profiles correspongding to the treatment gene expression profiles.
treatment	a matrix, is composed of gene expression profiles.

Details

The genes on the array are rank-ordered according to their differential expression relative to the control. First, control and treatment values less than a primary threshold value (quartile) were set to that threshold value. Finally, probe sets were ranked in descending order of d, where d is the ratio of the corresponding treatment-to-control values. For probe sets where d=1, a lower threshold was applied to the original difference values and a new treatment to control ratio (d') calculated. These probe sets were then sub-sorted in descending order of d.

Value

A matrix is composed of ranked lists, a ranked list represents the correspongding gene expression profiles.

Examples

```
if (require(GEOquery)){
  #treatment gene-expression profiles
  GSM118720 <- getGEO(filename=system.file("extdata/GSM118720.soft",package=
  "GeneExpressionSignature"))
  #control gene-expression profiles
  GSM118721 <- getGEO(filename=system.file("extdata/GSM118721.soft",package=
  "GeneExpressionSignature"))
  #data ranking according to the different expression values
  control <- as.matrix(as.numeric(Table(GSM118721)[,2]))
  treatment <- as.matrix(as.numeric(Table(GSM118720)[,2]))
  ranked_list <-getRLs(control,treatment)
}
```

RankMerging *Merging the ranker lists with the same labels of the biological states into a single list with the Iorio's method.*

Description

Merging the assay data according to phenotypic data of the input ExpressionSet. Each group of the ranked lists with the same phenotypic data is aggregated into a single list, return it as an ExpressionSet object.

Usage

```
RankMerging(exprSet, MergingDistance = c("Spearman", "Kendall"), weighted = TRUE)
```

Arguments

exprSet	an ExpressionSet object, each column of assay data represents a ranked list obtained by preprocessing the corresponding gene expression profile, and phenotypic data represents the short description (characteristics of gene expression profile, such as the drug type, the disease state) about the assay data.
MergingDistance	distance to be used which "measures" the similarity of ordered lists, the default is "Spearman"
weighted	there are tow rank merging approaches for two cases: if weighted=FALSE, all ranked list with the same biological state are treated equally important, a simple but useful method average ranking technique is selected; otherwise, weighted=TRUE, each individual ranked lists has its own ranked weights, this takes the iterative rank-aggregating algorithm, default is TRUE.

Details

The krubor function is used in the aggregating procedure. And the following methods are used in the implementation: a measure of the distance between two ranked lists (Spearman's Footrule), a method to merge two or more ranked lists the (Borda Merging Method), and a algorithm to obtain a single ranked list from a set of them in a hierarchical way (the Kruskal Algorithm). If choose Kendall as distance, the effectiveness of this function is certainly limited by the size of the merging problem.

See Also

SignatureDistance

Examples

```
#load the sample expressionSet
data(exampleSet)
```

```
## Merging each group of the ranked lists in the exampleSet with the same phenotypic data into a single PRL
MergingSet=RankMerging(exampleSet,"Spearman",weighted=TRUE)
```

ScoreGSEA *Compute pairwise distances between samples with method in package GSEA*

Description

Compute pairwise distances between sample according to their (Prototype Ranked List) PRL, a N x N distance matrix is generated by calling this function, N is the length of PRL.

Usage

```
ScoreGSEA(MergingSet, SignatureLength, ScoringDistance = c("avg", "max"), p.value = F)
```

Arguments

MergingSet	an ExpressionSet object. The assay data represents the PRLs of the samples, each column represents one PRL. The number of sample of this argument must be greater than 1, otherwise, this function is not meaningful.
SignatureLength	the length of "gene signature". In order to compute pairwise distances among samples, genes lists are ranked according to the gene expression ratio (fold change). And the "gene signature" includes the most up-regulated genes (near the top of the list) and the most down-regulated genes (near the bottom of the list).
ScoringDistance	the distance measurements between PRLs: the Average Enrichment Score Distance (avg), and the Maximum Enrichment Score Distance (max).
p.value	logical, if TRUE return a matrix of p.values of the distance matrix, default FALSE

Details

Once the PRL obtained for each sample, the distances between samples are calculated base on gene signature, including the expression of genes that seemed to consistently vary in response to the across different experimental conditions (e.g., different cell lines and different dosages). We take two distance measurements between PRLs: the Average Enrichment-Score Distance Davg=(TESx,y+TESy,x)/2, and the Maximum Enrichment-Score Distance Dmax=Min(TESx,y,TESy,x)/2.The avg is more stringent than max, where max is more sensitive to weak similarities, with lower precision but large recall.

Value

an distance-matrix, the max distance is more sensitive to weak similarities, providing a lower precision but a larger recall.

If p.value is set to TRUE, then a list is returned that consists of the distance matrix as well as their p.values, otherwise, without p.vlues in the result.

See Also

ScorePGSEA, SignatureDistance

Examples

```
# load the sample expressionSet
data(exampleSet)

# Merging each group of the ranked lists in the exampleSet with the same phenotypic data into a single PRL
MergingSet=RankMerging(exampleSet,"Spearman")

# get the distance matrix
ds=ScoreGSEA(MergingSet,250,"avg")
```

ScorePGSEA *Compute pairwise distances between samples with method in package PGSEA*

Description

Compute pairwise distances between sample according to their (Prototype Ranked List) PRL, get a N x N distance matrix is generated by calling this function , N is the length of PRL.

Usage

```
ScorePGSEA(MergingSet, SignatureLength, ScoringDistance = c("avg", "max"), p.value = F)
```

Arguments

MergingSet	an ExpressionSet object. The assay data represents the PRLs of the samples, each column represents one PRL. The number of sample must be greater than 1, oherwise, this function is not meaningful.
SignatureLength	the length of "gene signature". In order to compute pairwise distances among samples, genes lists are ranked according to the gene expression ratio (fold change). And the "gene signature" includes the most up-regulated genes (near the top of the list) and the most down-regulated genes (near the bottom of the list).
ScoringDistance	the distance measurements between PRLs: the Average Enrichment Score Distance (avg), or the Maximum Enrichment Score Distance (max).

p.value logical, if TRUE return a matrix of p.values of the distance matrix, default FALSE

Details

This function has the same function with ScoreGSEA, just with different methods.

See Also

ScoreGSEA, SignatureDistance

Examples

```
# load the sample expressionSet
data(exampleSet)

# Merging each group of the ranked lists in the exampleSet with the same phenotypic data into a single PRL
MergingSet=RankMerging(exampleSet,"Spearman")

# get the distance matrix
ds=ScorePGSEA(MergingSet,250, ScoringDistance="avg")
```

SignatureDistance *Compute pairwise distances comprehensively.*

Description

This function integrated the function for rank merging and distance scoring, we can do the rank merging and distance scoring simply with it.

Usage

```
SignatureDistance(exprSet, SignatureLength, MergingDistance = c("Spearman", "Kendall"), ScoringMethod
```

Arguments

exprSet an ExpressionSet object, each column of assay data represents a ranked list obtained by preprocessing the corresponding gene expression profile, and phenotypic data represents the short description (characteristics of gene expression profile, such as the drug type, the disease state) about the assay data.

SignatureLength the length of "gene signature". In order to compute pairwise distances among samples, genes lists are ranked according to the gene expression ratio (fold change). And the "gene signature" includes the most up-regulated genes (near the top of the list) and the most down-regulated genes (near the bottom of the list).

MergingDistance distance to be used which "measures" the similarity of ordered lists, Spearman or Kendall

ScoringMethod	method to be used to perform distance scoring, GSEA or PGSEA
ScoringDistance	the distance measurements between PRLs: the Average Enrichment Score Distance (avg), and the Maximum Enrichment Score Distance (max).
weighted	there are tow rank merging approaches for two cases: if weighted=FALSE, all ranked list with the same biological state are treated equally important, a simple but useful method average ranking technique is selected; otherwise, weighted=TRUE, each individual ranked lists has its own ranked weights, this takes the iterative rank-aggregating algorithm, default is TRUE.
...	additional arguments can be passed to the internal procedures

See Also

RankMerging,ScoreGSEA, ScorePGSEA

Examples

```
#load the sample expressionSet
data(exampleSet)

#distance scoring
SignatureDistance(exampleSet,SignatureLength=250,MergingDistance="Spearman", ScoringMethod="GSEA",ScoringDistar
```

Index