Package 'GSReg'

April 23, 2016

Version 1.4.0

Date 2014-04-08

Title Gene Set Regulation (GS-Reg)

Author Bahman Afsari <bahman@jhu.edu>, Elana J. Fertig
<ejfertig@jhmi.edu>

Maintainer Bahman Afsari <bahman@jhu.edu>

Depends R (>= 2.13.1)

Suggests GSBenchMark

Description A package for gene set analysis based on the variability of expressions. It implements DIfferential RAnk Conservation (DIRAC) and gene set Expression Variation Analysis (EVA) methods.

License GPL-2

biocViews GeneRegulation, Pathways, GeneExpression, GeneticVariability, GeneSetEnrichment

NeedsCompilation yes

R topics documented:

`GSReg-package` *A package for Gene Set Analysis based on the variability of gene expression in different phenotypes.*

Description

The GSReg package applies the analysis of variance among phenotypes for each gene set. Specially, the user can use Differential Rank Conservation (DIRAC) (Eddy et al. 2010) and a modified version which allows for efficient and easy p value calculation. Both DIRAC and its modified version are rank-based methods, i.e. they only consider the ordering of the expressions within the pathway.

GSReg package features

The package contains several utilities enabling to:

A) Prune Gene Sets based on the available genes in the expression data;

B) Calculate the DIRAC measure and p-value for it based on permutation test;

C) Calculate for a modified DIRAC method and a fast-efficient p-valuebased on U-Statistic theory;

Author(s)

Bahman Afsari <bahman.afsari@gmail.com>, Elana J. Fertig <ejfertig@jhmi.edu>

Source

http://www.ncbi.nlm.nih.gov/pubmed/20523739

References

Eddy et al., "Identifying tightly regulated and variably expressed networks by Differential Rank Conservation (DIRAC).", *PLoS Comp. Bio.*, 2010, **6**(5)

`GSReg.GeneSets.DIRAC` *Performs DIRAC for gene set analysis from the paper Eddy et al (2010).*

Description

GSReg.GeneSets.DIRAC performs DIRAC for gene set analysis from the paper Eddy et al (2010). In fact, the Null hypothesis is that the conservation index is not significantly different under two phenotypes. The function calculates the p-value using permutation test; hence, extremely low p-value cannot be reached.

Usage

```
GSReg.GeneSets.DIRAC(geneexpres, pathways, phenotypes, Nperm = 1000,minGeneNum=5)
```

Arguments

geneexpres	the matrix of gene expressions. The rownames must represent gene names and the columns represent samples. There must not be any missing values. Please use imputation or remove the genes with missing values.
pathways	a list containing pathway information. Each element represents a pathway as a character vector. The genes shown in the pathway must be present in geneexpres. Please prune the genes in pathways using GSReg.Prune before applying this function.
phenotypes	a binary factor containing the phenotypes for samples in geneexpres; hence, the column number of geneexpres and the length of phenotypes must be equal.
Nperm	The number of permutation tests required for p-value calculation. If Nperm==0 then no p-value is calculated.
minGeneNum	the minimum number of genes required in a pathway.

Value

IThe output is a list with three elements. Each element of the output list is a vector are named according to the pathway.

$mu1	a vector containing the variability in DIRAC sense (1- conservation indices in Eddy et al (2010) paper) for all pathways in phenotype == levels(phenotypes)[1].
$mu2	a vector containing the variability in DIRAC sense (1- conservation indices in Eddy et al (2010) paper) for all pathways in phenotype == levels(phenotypes)[2].
$pvalues	a vector containing p-values for each pathway. Low p-values means that the gene expressions have different orderings under different phenotypes.

Author(s)

Bahman Afsari

References

Eddy, James A., et al. "Identifying tightly regulated and variably expressed networks by Differential Rank Conservation (DIRAC)." PLoS computational biology 6.5 (2010): e1000792.

See Also

GSReg.GeneSet.VReg

Examples

```
library(GSBenchMark)
### loading and pruning the pathways
data(diracpathways)
### loading the data
data(leukemia_GSEA)
```

```
### extracting gene names
genenames = rownames(exprsdata);


### DIRAC analysis
DIRAna = GSReg.GeneSets.DIRAC(pathways=diracpathways,geneexpres=exprsdata,Nperm=10,phenotypes=phenotypes)
dysregulatedpathways = rbind(DIRAna$mu1[which(DIRAna$pvalues<0.05)],
DIRAna$mu2[which(DIRAna$pvalues<0.05)],DIRAna$pvalues[which(DIRAna$pvalues<0.05)]);
rownames(dysregulatedpathways)<-c("mu1","mu2","pvalues");
print(dysregulatedpathways[,1:5])
plot(x=dysregulatedpathways["mu1",],y=dysregulatedpathways["mu2",],
xlim=range(dysregulatedpathways[1:2,]),ylim=range(dysregulatedpathways[1:2,]))
lines(x=c(min(dysregulatedpathways[1:2,]),max(dysregulatedpathways[1:2,])),
y=c(min(dysregulatedpathways[1:2,]),max(dysregulatedpathways[1:2,])),type="l")
```

`GSReg.GeneSets.EVA` *Performs Gene Set Analysis using Expression Variation Analysis (EVA).*

Description

GSReg.GeneSets.EVA performs modified version DIRAC papers. Using a theoretical analysis, we can calculate p-value which makes extreme low p-values available.

Usage

```
GSReg.GeneSets.EVA(geneexpres, pathways, phenotypes,minGeneNum=5)
```

Arguments

`geneexpres`	the matrix of gene expressions. The rownames must represent gene names and the columns represent samples. There must not be any missing values. Please use imputation or remove the genes with missing values.
`pathways`	a list containing pathway information. Each element represents a pathway as a character vector. The genes shown in the pathway must be present in geneexpres. geneexpres must have numeric and finite numbers.
`phenotypes`	a binary factor containing the phenotypes for samples in geneexpres; hence, the column number of geneexpres and the length of phenotypes must be equal.
`minGeneNum`	the minimum number of genes required in a pathway.

Value

a list of analysis for all pathways.

`$E1`	the modified variance on the pathway within the samples from levels(phenotypes)[1].
`$E2`	the modified variance on the pathway within the samples from levels(phenotypes)[2].

$VarEta1	the estimation of the modified variance on the pathway within the samples from levels(phenotypes)[1].
$VarEta2	the estimation of the modified variance on the pathway within the samples from levels(phenotypes)[2].
$zscore	zscore for the modified variance.
$pvalues	theoretical p-value from the zscore.

Author(s)

Bahman Afsari

See Also

GSReg.GeneSets.DIRAC,cor

Examples

```
### loading and pruning the pathways
library(GSBenchMark)
data(diracpathways)
### loading the data
data(leukemia_GSEA)


### removing genes which contain not a number.
if(sum(apply(is.nan(exprsdata),1,sum)>0))
  exprsdata = exprsdata[-which(apply(is.nan(exprsdata),1,sum)>0),];

### extracting gene names
genenames = rownames(exprsdata);

### DIRAC analysis
VarAnKendallV = GSReg.GeneSets.EVA(geneexpres=exprsdata,
 pathways=diracpathways, phenotypes=as.factor(phenotypes))
E1 = sapply(VarAnKendallV,function(x) x$E1);
E2 = sapply(VarAnKendallV,function(x) x$E2);
Kpvalues = sapply(VarAnKendallV,function(x) x$pvalue);

dysregulatedpathways = rbind(E1[which(Kpvalues<0.05)],
E2[which(Kpvalues<0.05)],Kpvalues[which(Kpvalues<0.05)]);
rownames(dysregulatedpathways)<-c("E1","E2","pvalues");
print(dysregulatedpathways)
plot(x=dysregulatedpathways["E1",],y=dysregulatedpathways["E2",],
xlim=range(dysregulatedpathways[1:2,]),ylim=range(dysregulatedpathways[1:2,]))
lines(x=c(min(dysregulatedpathways[1:2,]),max(dysregulatedpathways[1:2,])),
y=c(min(dysregulatedpathways[1:2,]),max(dysregulatedpathways[1:2,])),type="l")
```

Index