Package 'SPIA'

April 10, 2015

Version 2.18.0

Date 2013-2-20

Title Signaling Pathway Impact Analysis (SPIA) using combined evidence of pathway over-representation and unusual signaling perturbations

Author Adi Laurentiu Tarca <atarca@med.wayne.edu>, Purvesh Kathri <purvesh@cs.wayne.edu> and Sorin Draghici <sorin@wayne.edu>

Depends R (>= 2.14.0), graphics, KEGGgraph

Suggests graph, Rgraphviz, hgu133plus2.db

Maintainer Adi Laurentiu Tarca <atarca@med.wayne.edu>

Description This package implements the Signaling Pathway Impact Analysis (SPIA) which uses the information form a list of differentially expressed genes and their log fold changes together with signaling pathways topology, in order to identify the pathways most relevant to the condition under the study.

License GPL (>= 2)

URL http://bioinformatics.oxfordjournals.org/cgi/reprint/btn577v1

Collate spia.R plotP.R combfunc.R getP2.R makeSPIAdata.R

Imports graphics

LazyLoad yes

biocViews Microarray, GraphAndNetwork

R topics documented:

`colorectalcancer` *Results from a microarray experiment comparing colorectal cancer samples and normal tissue samples.*

Description

The `colorectal` dataset consists: i) an named vector `DE_Colorectal` , which represents the $log2$ fold changes of the genes chosen as differentially expressed between colorectal cancer and normal samples based on data from Hong et al, 2007, using a $FDR=0.1$ and the universe of all Entrez gene IDs available on the array, `ALL_Colorectal`. These two vectors were obtained starting from the `top` dataframe which is the output from the topTable function of the limma package using the RMA processed gene expression data downloaded from GEE (GSE4107). The microarray platform used was Affymetrix HGU-133PLUS2.0.

Usage

```
data(colorectalcancer)
```

Source

Yi Hong and Kok Sun Ho and Kong Weng Eu and Peh Yean Cheah, A susceptibility gene set for early onset colorectal cancer that integrates diverse signaling pathways: implication for tumorigenesis, Clin Cancer Res, 2007, 13(4),1107-14.

`combfunc` *Combining two p-values using Fisher's product or normal inversion method*

Description

Combining two p-values using Fisher's product or normal inversion methods.

Usage

```
combfunc(p1=NULL,p2=NULL,combine="fisher")
```

Arguments

`p1`	A vector of probabilities.
`p2`	A vector of probabilities.
`combine`	A string with the name of the method to be used. Options include "fisher","norminv"

Details

Two vectors of p-values are combined into a vector of global p-values.

Value

A vector of p-values.

Author(s)

Adi Laurentiu Tarca <atarca@med.wayne.edu>, Purvesh Khatri, Sorin Draghici

References

Adi L. Tarca, Sorin Draghici, Purvesh Khatri, et. al, A Signaling Pathway Impact Analysis for Microarray Experiments, 2008, Bioinformatics, 2009, 25(1):75-82.

See Also

spia

Examples

```
# Examples use colorectal cancer dataset
p1=c(0.2,0.4,0.1)
p2=c(0.01,0.7,0.01)
pG=combfunc(p1,p2,combine="fisher")
pG=combfunc(p1,p2,combine="norminv")
```

---

makeSPIAdata *Process KGML files for spia analysis*

Description

This function processes KEGG xml files into a xxxSPIA.RData file needed for spia function.

Usage

```
makeSPIAdata(kgml.path="./hsa",organism="hsa",out.path=".")
```

Arguments

kgml.path	Character vector giving the location of the folder containing two or more KEGG xml files. See for e.g. http://www.genome.jp/kegg/pathway/hsa/hsa04010.html and click the Download KGML to get such files. Users that have a license to the KEGG ftp directory can copy all the xml files corresponding to a givne organism.
organism	A three letter character designating the organism. See a full list at ftp://ftp.genome.jp/pub/kegg/xml/organi
out.path	Directory where a "organism"SPIA.RData file will be saved. If left to null, it will will try to save the file in th extdata folder of the SPIA library.

Author(s)

Adi Laurentiu Tarca <atarca@med.wayne.edu>

See Also

spia

Examples

```
library(SPIA)
data(colorectalcancer)
makeSPIAdata(kgml.path=system.file("extdata/keggxml/hsa",package="SPIA"),organism="hsa",out.path="./")
res<-spia(de=DE_Colorectal, all=ALL_Colorectal, organism="hsa",data.dir="./")
res[,-12]
```

---

plotP *SPIA two-way evidence plot*

Description

Plots each pathway as a point, using the over-representation p-value, pNDE, and perturbations accumulation p-value, pPERT, as coordinates. In addition the regions where FDR and FWER adjusted pG values are less than the specified threshold are plotted. The function determines automatically which method (fisher or norminv) was used to combine the two p-values into pG, and plots the regions described above accordingly.

Usage

```
plotP(x,threshold=0.05)
```

Arguments

x	A data frame produced by spia function.
threshold	A numerical value between 0 and 1 to be used as significance threshold in inferring pathway significance.

Details

In this plot each pathway is a point and the coordinates are the log of pNDE (using a hypergeometric model) and the p-value from perturbations, pPERT. The oblique lines in the plot show the significance regions based on the combined evidence.

Value

This function does not return any value. It only generates a plot.

Author(s)

Adi Laurentiu Tarca <atarca@med.wayne.edu>, Purvesh Khatri, Sorin Draghici

References

Adi L. Tarca, Sorin Draghici, Purvesh Khatri, et. al, A Signaling Pathway Impact Analysis for Microarray Experiments, 2008, Bioinformatics, 2009, 25(1):75-82.

See Also

spia

Examples

```
# Examples use colorectal cancer dataset
data(colorectalcancer)

# pathway analysis based on combined evidence of ORA and perturbations
# use nB=2000 or larger for more accurate results
res<-spia(de=DE_Colorectal, all=ALL_Colorectal, organism="hsa",nB=200,plots=FALSE,verbose=TRUE,beta=NULL,combin

#Generate the evidence plot
plotP(res,threshold=0.1)

res<-spia(de=DE_Colorectal, all=ALL_Colorectal, organism="hsa",nB=200,plots=FALSE,verbose=TRUE,beta=NULL,combin

#Generate the evidence plot
plotP(res,threshold=0.1)
```

`spia` *Signaling Pathway Impact Analysis (SPIA) based on over-representation and signaling perturbations accumulation*

Description

This function implements the SPIA algorithm to analyze KEGG signaling pathways.

Usage

```
spia(de=NULL,all=NULL,organism="hsa",data.dir=NULL,pathids=NULL,nB=2000,plots=FALSE,verbose=TRUE,be
```

Arguments

de	A named vector containing log2 fold-changes of the differentially expressed genes. The names of this numeric vector are Entrez gene IDs.
all	A vector with the Entrez IDs in the reference set. If the data was obtained from a microarray experiment, this set will contain all genes present on the specific array used for the experiment. This vector should contain all names of the de argument.
organism	A three letter character designating the organism. See a full list at ftp://ftp.genome.jp/pub/kegg/xml/organi
data.dir	Location of the "organism"SPIA.RData file containing the pathways data generated with makeSPIAdata. If set to NULL will look for this file in the extdata folder of the SPIA library.
pathids	A character vector with the names of the pathways to be analyzed. If left NULL all pathways available will be tested.
nB	Number of bootstrap iterations used to compute the P PERT value. Should be larger than 100. A recommended value is 2000.
plots	If set to TRUE, the function plots the gene perturbation accumulation vs log2 fold change for every gene on each pathway. The null distribution of the total net accumulations from which PPERT is computed, is plotted as well. The figures are sent to the SPIAPerturbationPlots.pdf file in the current directory.
verbose	If set to TRUE, displays the number of pathways already analyzed.
beta	Weights to be assigned to each type of gene/protein relation type. It should be a named numeric vector of length 23, whose names must be: c("activation","compound","binding/as If set to null, beta will be by default chosen as: c(1,0,0,1,-1,1,0,0,-1,-1,0,0,1,0,1,-1,0,1,-1,-1,0,0,0).
combine	Method used to combine the two types of p-values. If set to "fisher" it will use Fisher's method. If set to "norminv" it will use the normal inversion method.

Details

See cited documents for more details.

Value

A data frame containing the ranked pathways and various statistics: pSize is the number of genes on the pathway; NDE is the number of DE genes per pathway; tA is the observed total preturbation accumulation in the pathway; pNDE is the probability to observe at least NDE genes on the pathway using a hypergeometric model; pPERT is the probability to observe a total accumulation more extreme than tA only by chance; pG is the p-value obtained by combining pNDE and pPERT; pGFdr and pGFWER are the False Discovery Rate and respectively Bonferroni adjusted global p-values; and the Status gives the direction in which the pathway is perturbed (activated or inhibited). KEGGLINK gives a web link to the KEGG website that displays the pathway image with the differentially expressed genes highlighted in red.

Author(s)

Adi Laurentiu Tarca <atarca@med.wayne.edu>, Purvesh Khatri, Sorin Draghici

References

Adi L. Tarca, Sorin Draghici, Purvesh Khatri, et. al, A Signaling Pathway Impact Analysis for Microarray Experiments, 2008, Bioinformatics, 2009, 25(1):75-82.

Purvesh Khatri, Sorin Draghici, Adi L. Tarca, Sonia S. Hassan, Roberto Romero. A system biology approach for the steady-state analysis of gene signaling networks. Progress in Pattern Recognition, Image Analysis and Applications, Lecture Notes in Computer Science. 4756:32-41, November 2007.

Draghici, S., Khatri, P., Tarca, A.L., Amin, K., Done, A., Voichita, C., Georgescu, C., Romero, R.: A systems biology approach for pathway level analysis. Genome Research, 17, 2007.

See Also

plotP

Examples

```
# Example using a colorectal cancer dataset obtained using Affymetrix geneChip technology (GEE GSE4107).
# Suppose that proper preprocessing was performed and a two group moderated t-test was applied. The topTable
# result from limma package for this data set is called "top".
#The following lines will annotate each probeset to an entrez ID identifier, will keep the most significant probeset
#gene ID and retain those with FDR<0.05 as differentially expressed.
#You can run these lines if hgu133plus2.db package is available

#data(colorectalcancer)
#x <- hgu133plus2ENTREZID
#top$ENTREZ<-unlist(as.list(x[top$ID]))
#top<-top[!is.na(top$ENTREZ),]
#top<-top[!duplicated(top$ENTREZ),]
#tg1<-top[top$adj.P.Val<0.1,]
#DE_Colorectal=tg1$logFC
#names(DE_Colorectal)<-as.vector(tg1$ENTREZ)
#ALL_Colorectal=top$ENTREZ

data(colorectalcancer)

# pathway analysis using SPIA; # use nB=2000 or higher for more accurate results
#uses older version of KEGG signalimng pathways graphs
res<-spia(de=DE_Colorectal, all=ALL_Colorectal, organism="hsa",beta=NULL,nB=2000,plots=FALSE, verbose=TRUE,comb
res
# Create the evidence plot
plotP(res)

#now combine pNDE and pPERT using the normal inversion method without running spia function again
res$pG=combfunc(res$pNDE,res$pPERT,combine="norminv")
res$pGFdr=p.adjust(res$pG,"fdr")
res$pGFWER=p.adjust(res$pG,"bonferroni")
plotP(res,threshold=0.05)
```

```
#highlight the colorectal cancer pathway in green
points(I(-log(pPERT))~I(-log(pNDE)),data=res[res$ID=="05210",],col="green",pch=19,cex=1.5)

#run SPIA using pathways data generated from (up-to-date) xml files that you can obtain from
#KEGG ftp or by downloading  them from each pathways web page:
# e.g. go to http://www.genome.jp/kegg/pathway/hsa/hsa04010.html  and click on DOwnload KGML
#to get the xml file for pathway 4010

makeSPIAdata(kgml.path=system.file("extdata/keggxml/hsa",package="SPIA"),organism="hsa",out.path="./")

res<-spia(de=DE_Colorectal, all=ALL_Colorectal, organism="hsa",data.dir="./")
res
```

`Vessels` *Results from a microarray expriment comparing umbilical veins and arteries tissues*

Description

The `Vessels` dataset consists an named vector `DE_Vessels` , which represents the log2 fold changes of the genes chosen as differentially expressed between ubilical veins and arteries tissue (Kim et al, 2008), and the universe of all Entrez gene IDs available on the array, `ALL_Vessels`. The microarray platform used was Illumina's Human-6 v2 expression BeadChip.

Usage

```
data(Vessels)
```

Source

These data was produced at the Perinatology Research Branch, of Wayne State University (Detroit), and accompanies the publication:
Kim JS, Romero R, Tarca A, Lajeunesse C, Han YM, Kim MJ, Suh YL, Draghici S, Mittal P, Gotsch F, Kusanovic JP, Hassan S, Kim CJ, Gene expression profiling demonstrates a novel role for fetal fibrocytes and the umbilical vessels in human fetoplacental development, J Cell Mol Med, 2008, PMID: 18298660.

Index