SPIA

November 11, 2009

R topics documented:

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

Description

The `colorectal` dataset consists: 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, 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 GEO (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.

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.

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

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

SPIA-internal *Internal SPIA functions*

Description

Internal SPIA functions. `combfunc` combines two p-values into a global significance p-value. `getP2` computes the product of two independent p-values that corresponds to a given global p-value.

Usage

```
combfunc(p1,p2)
getP2(t)
```

Details

These are not to be called directly by the user.

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

Description

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

Usage

```
spia(de=NULL,all=NULL,organism="hsa",nB=2000,plots=FALSE,verbose=TRUE,beta=NULL)
```

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/or .
`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","b "indirect","inhibition_phosphorylation","dephosphorylation_inhibiti "state","activation_indirect","inhibition_ubiquination","ubiquinati "binding/association_phosphorylation","dissociation_phosphorylation` 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).

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).

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
# Suppose that proper preprocessing was performed and a two group moderated t-test was ap
# 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 th
#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$adj.P.Val<0.05,]
#DE_Colorectal=tg1$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
res<-spia(de=DE_Colorectal, all=ALL_Colorectal, organism="hsa",beta=NULL,nB=200,plots=FAL
res
# Create the evidence plot
plotP(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