

# Package ‘SPIA’

May 16, 2024

**Version** 2.57.0

**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 from 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** file LICENSE

**License\_restricts\_use** yes

**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

**git\_url** <https://git.bioconductor.org/packages/SPIA>

**git\_branch** devel

**git\_last\_commit** 40a790f

**git\_last\_commit\_date** 2024-04-30

**Repository** Bioconductor 3.20

**Date/Publication** 2024-05-15

## Contents

colorectalancer . . . . .	2
combfunc . . . . .	2
makeSPIAdata . . . . .	3
plotP . . . . .	4
spia . . . . .	5
SPIA-internal . . . . .	8
Vessels . . . . .	8

<b>Index</b>	<b>10</b>
--------------	-----------

---

colorectalancer	<i>Results from a microarray experiment comparing colorectal cancer samples and normal tissue samples.</i>
-----------------	--

---

### Description

The colorectal dataset consists: i) an named vector `DE_Colorectal` , which represents the  $\log_2$  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(colorectalancer)
```

### 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	<i>Combining two p-values using Fisher's product or normal inversion method</i>
----------	---

---

### 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. <a href="http://www.genome.jp/kegg/pathway/hsa/hsa04010.html">http://www.genome.jp/kegg/pathway/hsa/hsa04010.html</a> 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 given organism.
organism	A three letter character designating the organism. See a full list at <a href="ftp://ftp.genome.jp/pub/kegg/xml/organism">ftp://ftp.genome.jp/pub/kegg/xml/organism</a>
out.path	Directory where a "organism".SPIA.RData file will be saved. If left to null, it will try to save the file in the 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 <a href="#">spia</a> 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.



**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$pG,"fdr")
res$pFWER=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 pathway's web page:
# e.g. go to http://www.genome.jp/kegg/pathway/hsa/hsa04010.html and click on D0wnload 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

```

---

SPIA-internal

*Internal SPIA functions*

---

### Description

Internal SPIA functions. getP2 given a pG returns two equal p-values such as  $\text{combfunc}(p1,p2)=pG$ .

### Usage

```
getP2(pG,combine="fisher")
```

### Details

These are not to be called directly by the user.

---

Vessels

*Results from a microarray experiment 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 umbilical 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

- \* **datasets**
  - colorectalancer, 2
  - Vessels, 8
- \* **internal**
  - SPIA-internal, 8
- \* **methods**
  - combfunc, 2
  - makeSPIAdata, 3
  - plotP, 4
  - spia, 5
- \* **nonparametric**
  - combfunc, 2
  - makeSPIAdata, 3
  - plotP, 4
  - spia, 5

ALL\_Colorectal (colorectalancer), 2  
ALL\_Vessels (Vessels), 8

colorectalancer, 2  
combfunc, 2

DE\_Colorectal (colorectalancer), 2  
DE\_Vessels (Vessels), 8

getP2 (SPIA-internal), 8

makeSPIAdata, 3

plotP, 4, 7

spia, 3–5, 5  
SPIA-internal, 8

top (colorectalancer), 2

Vessels, 8