Package 'MPINet'

February 19, 2015

Version 1.0

Title The package can implement the network-based metabolite pathway identification of pathways.

Author Yanjun Xu, Chunquan Li and Xia Li

Maintainer Yanjun Xu <tonghua605@163.com>

Description (1) Our system provides a network-based strategies for metabolite pathway identification.(2) The MPINet can support the identification of pathways using Hypergeometric test based on metabolite set. (3)MPINet can support pathways from multiple databases.

Depends R (>= 2.15.2),BiasedUrn,mgcv

Collate getPSS.R performpcls.R identifypathway.R GetExampleData.R getEnvironmentData.R

LazyData Yes

License GPL (>= 2)

biocViews Statistics, Annotation, Pathways, metabolite, Networks

NeedsCompilation no

Repository CRAN

Date/Publication 2013-07-28 08:30:30

R topics documented:

ExampleData *The example data in the environment variable of the system*

Description

The example data in the environment variable of the system.

Format

An environment variable

Details

The example data includes the variable `prostateTMrisk`, `diabetes1` and `diabetes2`. The `prostateTMrisk` is the differential metabolites from the study of Sreekumar et al., the `diabetes1` from text mining and the `diabetes2` is the differential metabolite set from the study of Suhre et al.. These three example datasets can be obtained by the function `GetExampleData`.

Author(s)

Yanjun Xu <tonghua605@163.com>, Chunquan Li <lcqbio@aliyun.com.cn> and Xia Li <lixia@hrbmu.edu.cn>

References

Sreekumar, A., Poisson, L.M., Rajendiran, T.M., Khan, A.P., Cao, Q., Yu, J., Laxman, B., Mehra, R., Lonigro, R.J., Li, Y. et al. (2009) Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature, 457, 910-914.

Suhre, K., Meisinger, C., Doring, A., Altmaier, E., Belcredi, P., Gieger, C., Chang, D., Milburn, M.V., Gall, W.E., Weinberger, K.M. et al. (2010) Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. PLoS ONE, 5, e13953.

GetExampleData *Get the example data set*

Description

Get the example data set.

Usage

```
GetExampleData(dataset)
```

Arguments

`dataset`	A character string, must be one of "prostate" (default), "diabetes1" and "diabetes2".

Details

The example data are obtained from the environment variable MPINetData.

Value

A character vector of interesting metabolites, for each element is a pubchem CID

Author(s)

Yanjun Xu <tonghua605@163.com>, Chunquan Li <lcqbio@aliyun.com.cn> and Xia Li <lixia@hrbmu.edu.cn>

References

Sreekumar, A., Poisson, L.M., Rajendiran, T.M., Khan, A.P., Cao, Q., Yu, J., Laxman, B., Mehra, R., Lonigro, R.J., Li, Y. et al. (2009) Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature, 457, 910-914.

Suhre, K., Meisinger, C., Doring, A., Altmaier, E., Belcredi, P., Gieger, C., Chang, D., Milburn, M.V., Gall, W.E., Weinberger, K.M. et al. (2010) Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. PLoS ONE, 5, e13953.

Examples

```
## Not run:

#obtain the metastatic prosate cancer data set
Exampledata<-GetExampleData(dataset="prostate")
head(Exampledata)


## End(Not run)
```

---

getPSS *Evaluate the CGNB score of metabolites*

Description

Integrate the non-equivalence scores and the initial bias scores of metabolites by the monotonic spline model.

Usage

```
getPSS(riskmeta,plot=TRUE,binsize=400)
```

Arguments

riskmeta	A character vector of interesting metabolites, for each element is a pubchem CID.
plot	A logical. If TRUE the fit line obtained by the monotonic spline model will be plotted.
binsize	plot the fraction of differential metabolites in bins of this size.

Details

This function is used to integrate the non-equivalence of metabolites and the character of differential metabolites.The binsize can be assigned according to the best visualization.

Value

A data frame with 4 columns including "riskmeta", "meanstrvalue", "pss" and "CGNB". Each row correspond a metabolite pubchem CID. "riskmeta" indicates whether the metabolite is in the interesting set (with "1" is in and "0" is not in)."meanstrvalue" is the mean SOC value of the metabolite. "pss" is the score value obtained by the monotonic spline model. "CGNB" is the CGNB score of metabolite which is calculated as 1 substract the score value obtained by monotonic spline model. This score is used to calculate pathway weight in the subsequent pathway analysis.

Author(s)

Yanjun Xu <tonghua605@163.com>, Chunquan Li <lcqbio@aliyun.com.cn> and Xia Li <lixia@hrbmu.edu.cn>

References

Young, M.D., Wakefield, M.J., Smyth, G.K. and Oshlack, A. (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol, 11, R14.

Examples

```
## Not run:

########################################


##########get example data
risk<-GetExampleData(dataset="prostate")

###########calculate the CGNB score
pss<-getPSS(risk ,plot=F)
CGNBscore<-pss[,"CGNB"]
names(CGNBscore)<-rownames(pss)
##########print the CGNB score of some metabolites to screen
head(CGNBscore)
#identify dysregulated pathways
anncpdpre<-identifypathway(risk,pss,pathType="KEGG",method="MPINet",annlim=1,bglim=6)
#convert ann to data.frame
result<-printGraph(anncpdpre,pathType="KEGG",method="MPINet")
```

```
head(result)




## End(Not run)
```

---

identifypathway *A novel pathway identification approach based on metabolite set*

Description

Identify pathways via global weighted human metabolite network, which considering both the global non-equivalence of metabolites in pathway and the bias existing in metabonomic experiment technology.

Usage

```
identifypathway(componentList,PSS,pathType="KEGG",method="MPINet",weightnum=6,
                backgroundcid=Getenvir("getBackground"),annlim=1,bglim=6,
                order="pvalue",decreasing=FALSE)
```

Arguments

componentList	A character vector of interesting metabolites, for each element is a pubchem CID.
PSS	A data frame, which is obtained by the function getPSS.
pathType	A character string vector specifying the pathway source,must be some of the elements in the vector c("KEGG", "consensusPath", "PharmGKB", "SMPDB", "Wikipathways", "PID", "Reactome", "INOH", "BioCarta", "HumanCyc", "EHMN"). The default is KEGG.
method	A character string specifying the pathway identification method, must be one of "MPINet"(default) and "Hyper".
weightnum	A value. The power of the relative weight of pathway. The default value is 6.
backgroundcid	A character vector of the background metabolites, which is used to identify the statistically significant pathways.
annlim	An integer. Only use pathways annotated at least this number of metabolites. The default value is 1.
bglim	An integer. Only use pathways containing at least this number of metabolites. The default value is 6.
order	A character string. Should be one of "pvalue" and "fdr".
decreasing	A logical. Should the sort order be increasing or decreasing?

Details

The function can annotate a set of metabolites to pathways and identify the statistically significantly enriched pathways. The argument `method` should be one of "MPINet" and "Hyper".When the "MPINet" is specified, MPINet method which is considers both the global non-equivalence of metabolites in pathway and bias existing in metabonomic experiment technology is used. When the "Hyper" method is selected, the Hypergeometric test is used. If users don't set the values of the argument `backgroundcid`, the human background metabolites will be obtained from our default data set which contains 4994 metabolites and selected from five databases including MSEA, HMDB, SMPDB, KEGG and Reactome. Note that the argument `weightnum` can be assigned according to the bias level, which can be evaluated through the `getPSS` function by set the `plot` argument as "TRUE". If the plot line is more closer to right up, the `weightnum` should be assigned a higher value.

Value

A list. Each element of the list is another list. (i)If the argument `method` is "MPINet", it includes the following elements: 'pathwayName', 'annComponentList', 'annComponentNumber', 'annBgComponentList', 'annBgNumber', 'componentNumber', 'bgNumber', 'pvalue', 'fdr', 'InWeight', 'weight', 'anncompinNetworkNum', 'anncompinNetworkList', 'riskcompinNetworkNum', 'riskcompinNetworkList'. They correspond to pathway name, the submitted metabolites annotated to a pathway, numbers of submitted metabolites annotated to a pathway, the background metabolites annotated to a pathway, numbers of background metabolites annotated to a pathway, numbers of submitted metabolites, numbers of background metabolitess, p-value of the Wallenius' noncentral hypergeometric test, Benjamini-Hochberg fdr values, the mean score value of metabolites in pathway, the final weight of pathway, numbers of the submitted metabolites annotated to a pathway and in the global human metabolite network, the submitted metabolites annotated to a pathway and in the global human metabolite network, numbers of submitted metabolites in the global human metabolite network, submitted metabolites in the global human metabolite network. When the argument `pathType` is "KEGG", the 'pathwayId' element is also included, which is the pathway identifier in KEGG. When the argument `pathType` is not "KEGG", the 'pathsource' element is also included, which stands for the source of pathway. (ii)If the argument `method` is "Hyper", it includes the same elements as (i), but not includes the following elements: 'InWeight', 'weight', 'anncompinNetworkNum', 'anncompinNetworkList', 'riskcompinNetworkNum', 'riskcompinNetworkList'. To save the results, the `list` can be converted to the `data.frame` by the function `printGraph`.

Note that `componentList` submitted by users must be a 'character' vector.

Author(s)

Yanjun Xu <tonghua605@163.com>, Chunquan Li <lcqbio@aliyun.com.cn> and Xia Li <lixia@hrbmu.edu.cn>

References

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 102, 15545-15550.

Li, C., Li, X., Miao, Y., Wang, Q., Jiang, W., Xu, C., Li, J., Han, J., Zhang, F., Gong, B. et al. (2009) SubpathwayMiner: a software package for flexible identification of pathways. Nucleic Acids Res, 37, e131.

Examples

```
## Not run:

#####identify pathways related with metastatic prostate cancer###########
#example 1
#get example data
#### get the metastatic prostate cancer interesting metabolite data set
risk<-GetExampleData(dataset="prostate")
#### integrate the global non-equivalence of metabolites and the character of
#### differential metabolites by the monotonic spline model
pss<-getPSS(risk)

#identify pathways
anncpdpre<-identifypathway(risk,pss,pathType="KEGG",method="MPINet",annlim=1,bglim=6)
#convert ann to data.frame
result<-printGraph(anncpdpre,pathType="KEGG",method="MPINet")
#print part of the results to screen
head(result)

##write the results to tab delimited file.
write.table(result,file="result.txt",row.names=FALSE,sep="\t")

result1<-printGraph(anncpdpre,pathType="KEGG",method="MPINet",detail=TRUE)
##write the results to tab delimited file.
write.table(result1,file="result1.txt",row.names=FALSE,sep="\t")


#example 2
#get example data from file
risk<-read.table(paste(system.file(package="MPINet"),"/localdata/prostate.txt",sep=""),
header=F,sep="\t","\"")

####convert the data to a character vector
risk<-as.character(risk[[1]])

pss<-getPSS(risk)

#identify pathways
anncpdpre<-identifypathway(risk,pss,pathType="KEGG",method="MPINet",annlim=1,bglim=6)
#convert ann to data.frame
result<-printGraph(anncpdpre,pathType="KEGG",method="MPINet")
#print part of the results to screen
head(result)

#example 3
#get example data
#### get the metastatic prostate cancer interesting metabolite data set
risk<-GetExampleData(dataset="prostate")
```

```
pss<-getPSS(risk)

#identify dysregulated Reactome and KEGG pathways
anncpdpre<-identifypathway(risk,pss,pathType=c("KEGG","Reactome"),
method="MPINet",annlim=1,bglim=6)
#convert ann to data.frame
result<-printGraph(anncpdpre,pathType=c("KEGG","Reactome"),method="MPINet")
#print part of the results to screen
head(result)



## End(Not run)
```

`MPINetData` *The variables in the environment variable* `MPINetData` *of the system*

Description

The variables in the environment variable `MPINetData` of the system.

Format

An environment variable

Author(s)

Yanjun Xu <tonghua605@163.com>, Chunquan Li <lcqbio@aliyun.com.cn> and Xia Li <lixia@hrbmu.edu.cn>

`printGraph` *Print the results of identification*

Description

Print the identification results of MPINet.

Usage

```
printGraph(ann,detail=FALSE,method="MPINet",pathType="KEGG")
```

Arguments

ann	A list. The results returned from the function identifypathway.
detail	A logical. If true, metabolite lists from the function identifypathway are converted into strings, which are used to display and write results with metabolites.
method	A character string. which is the argument method used in the function identifypathway.
pathType	A character string vector. which is the argument pathType used in the function identifypathway.

Details

The function can convert the results of identification to data.frame.

Note that the argument method and pathType should be assigned the same as the function identifypathway.

Value

A data.frame of the identification results. (i)If the argument method is "MPINet", it includes the following elements: 'pathwayName', 'annComponentRatio', 'annBgRatio', 'weight', 'pvalue', 'fdr', 'annComponentList', 'annBgComponentList', 'annComponentinNetRatio', 'anncompinNetworkList', 'riskcompinNetworkList'. The 'annComponentRatio' is the ratio of the annotated metabolites. For example, 30/1000 means that 30 metabolites in 1000 interesting metabolites are annotated in this pathway. The 'annBgRatio' is the ratio of background metabolites, for example, 10/4994 means that 10 of the 4994 human metabolites in the background are annotated in this pathway. The 'annComponentinNetRatio' indicates the ratio of annotated metabolites in the global human metabolite network. The 'annComponentList' and 'annBgComponentList' are the annotated metabolites and the annotated background metabolites. The 'anncompinNetworkList' and 'riskcompinNetworkList' are the annotated metabolites in network and the interesting metabolites in network. (ii)If the argument method is "Hyper", it includes the following elements: 'pathwayName', 'annComponentRatio', 'annBgRatio', 'pvalue', 'fdr', 'annComponentList', 'annBgComponentList'. When the argument pathType is "KEGG", the 'pathwayID' is included. When the the argument pathType is not "KEGG", the 'pathsource' is included. Detailed information is provided in the function identifypathway.

Author(s)

Yanjun Xu <tonghua605@163.com>, Chunquan Li <lcqbio@aliyun.com.cn> and Xia Li <lixia@hrbmu.edu.cn>

See Also

identifypathway

Examples

```
## Not run:

#######################################

#get example data
#### get the type 2 diabetes data set 1
```

```
risk<-GetExampleData(dataset="diabetes1")
#### integrate the non-equivalence of metabolites and the character of
#### differential metabolites by the monotonic spline model
pss<-getPSS(risk,plot=TRUE)
#identify dysregulated pathways
anncpdpre<-identifypathway(risk,pss,pathType="KEGG",method="MPINet",annlim=1,bglim=6)
#convert ann to data.frame
result<-printGraph(anncpdpre,pathType="KEGG",method="MPINet")
#print part of the results to screen
head(result)

result1<-printGraph(anncpdpre,pathType="KEGG",method="MPINet",detail=TRUE)
#print part of the results to screen
head(result1)







## End(Not run)
```

Index