

Package ‘DBGSA’

February 19, 2015

Type Package

Title methods of distance-based gene set functional enrichment analysis.

Version 1.2

Date 2011-12-29

Author Li Jin, Huang Meilin

Maintainer Li Jin <lijin@hrbmu.edu.cn>

Data 2011-05-17

Description This package provides methods and examples to support a method of Gene Set Analysis (GSA). DBGSA is a novel distance-based gene set enrichment analysis method. We consider that, the distance between 2 groups with different phenotype by focusing on the gene expression should be larger, if a certain gene functional set was significantly associated with a particular phenotype.

License GPL-3

Depends R (>= 2.12.2), fdrtool(>= 1.2.6)

BioViews bioinformatics, Microarray

Collate avelinkdis.R centdis.R distable.R randdis.R valuep.R
functionall.R

lazy-loading yes

Repository CRAN

Date/Publication 2012-01-16 07:32:40

NeedsCompilation no

R topics documented:

DBGSA-package	2
afExp	3
avelinkdis	3
b1	4

centdis	4
distable	5
distance	6
functionall	6
rand	8
randdis	8
valuep	9

Index	11
--------------	-----------

DBGSA-package	<i>methods of distance-based gene set functional enrichment analysis</i>
---------------	--

Description

DBGSA package provides a new method to evaluate the possibilities of the enrichment between different phenotypes. DBGSA is a novel distance-based gene set enrichment analysis method. We consider that, the distance between 2 groups with different phenotype by focusing on the gene expression should be larger, if a certain gene functional set was significantly associated with a particular phenotype.

Details

Package:	DBGSA
Type:	Package
Version:	1.1
Depends:	R(>=2.12.2),methods,fdrtool
Suggest:	ALL
Date:	2011-05-17
License:	GPL-3
LazyLoad:	yes
BiocViews:	bioinformatics

Author(s)

Li Jin, Huang Meilin Maintainer:Li Jin<lijin@hrbmu.edu.cn>

References

Lu Tian, Steven A. Greenberg, Sek Won Kong, Josiah Altschuler, Discovering statistically significant pathways in expression profiling studies, PNAS 2005 102;13544-13549

afExp

Gene expression profile

Description

A matrix of data specifying for the gene set expression profile, each row presents a expression profile of a specific gene, each column represents a sample.

Usage

```
data(afExp)
```

Format

These are both gene expression profiles, each row represents a gene, each column represents a sample

Source

They are based on the data obtained as followed, by deleting the first row and first column <http://www.broadinstitute.org/gsea/>

Examples

```
data(afExp)
```

avelinkdis

Computing the distance by group average method

Description

A function which is used to calculate the distances between case and control group by method of group average method

Usage

```
avelinkdis(C,num,Meth)
```

Arguments

C	An input matrix which contains the gene profiles
num	An integer indicating the number of case group
Meth	A character string indicates which method to be used to compute the distances between genes, euclidean and Manhattan are available to choose

Value

A matrix containing the distances

Author(s)

Li Jin, Huang Meilin

b1	<i>Gene set function expression profile</i>
----	---

Description

This is a data file containing three items, gene function label, gene name, gene expression profile, a function label may contain many genes, each gene contains many expression profiles, each row has a gene function label, a gene name and the gene's expression profile, each column represents a sample.

Source

These data are obtained from original data by some operations.

Examples

```
filename=system.file("data", "b1.txt", package="DBGSA")
```

centdis	<i>Computing the distance by gravity method</i>
---------	---

Description

A function which is used to calculate the distances between case and control group by method of gravity method

Usage

```
centdis(C,num,Meth)
```

Arguments

C	An input matrix which contains the gene profiles
num	An integer indicating the number of case group
Meth	A character string indicates which method to be used to compute the distances between genes, euclidean and Manhattan are available to choose

Value

A matrix containing the distances

Author(s)

Li Jin, Huang Meilin

distable	<i>function to compute the specific distances between different phenotypes</i>
----------	--

Description

function which is used to calculate the required distances according to the given math method, such as euclidean, then the result of the distances will be returned in a text

Usage

```
distable(fd,num,setdis,Meth,resultname)
```

Arguments

fd	Character string represents the name of a connection of the file to load, each row contains three items, Class labels of gene function, gene name and gene expression profile, each column represents the information of a class labels of gene function
num	An integer indicating the number of the case group
setdis	A character string indicating which method to be used to compute the distances between case group and control group, avelinkdis or centdis is the choice to choose
Meth	A character string indicates which method to be used to compute the distances between genes, euclidean and Manhattan is available to choose
resultname	Character string which represents the name of the output file

Value

A text containing the distances.

Author(s)

Li Jin, Huang Meilin

Examples

```
## Not run:

##compute the distances and put them in the text named juli, the input file named "b1.txt"
##before using "b1.txt" data, we need to find "b1.txt" in the data file and put it in the work space of R
filename=system.file("data","b1.txt",package="DBGSA")
distable(filename,10,avelinkdis,"euclidean","juli.txt")

## End(Not run)
```

distance	<i>distances between case and control group</i>
----------	---

Description

This is a data file that has three columns: gene function label, the number of gene the former gene function label contains, distances between case and control group in this gene function lable, each row of this matrix represents the information of a gene function label.

Usage

```
data(distance)
```

Source

these data are obtained by the distable function, and then delete the first row

Examples

```
data(distance)
```

functionall	<i>P values of distance-based gene set enrichment analysis</i>
-------------	--

Description

Function which cooperates all the functions claimed before, by using this function and the input data, we may get the p-values directly instead of running the functions step by step

Usage

```
functionall(fd,num,setdis,Meth,resultname,original,randgenenum,randtime,randname,presult)
```

Arguments

fd	Character string represents the name of a connection of the file to load, each row contains three items, Class labels of gene function, gene name and gene expression profile, each column represents the information of a class labels of gene function
num	An integer indicating the number of the case group
setdis	A character string indicating which method to be used to compute the distances between case group and control group, avelinkdis or centdis is the choice to choose
Meth	A character string indicates which method to be used to compute the distances between genes, euclidean and Manhattan is available to choose
resultname	Character string which represents the name of the output file which contains the distances between different label
original	Another input matrix representing the gene expression profile
randgenenum	An integer indicating the number of gene to be permuted, the number of it has to be larger than the gene number that any gene function label has.
randtime	An integer represents the number of permutation
randname	Character string representing the name of the output file which contains the result of permutation
result	Character string representing the name of the output file which contains the result of p values

Value

A text containing the p values

Author(s)

Li Jin, Huang Meilin

See Also

[distable](#), [randdis](#), [valuep](#)

Examples

```
## Not run:  
data(afExp)  
##before running this funtion, we need to find out "b1.txt" data file, and put it into the R workspace.  
functionall(system.file("data", "b1.txt", package="DBGSA"),10,avelinkdis,"euclidean","distance",afExp,2000,10,"r  
  
## End(Not run)
```

rand	<i>distances of gene expression obtained by gene resampling</i>
------	---

Description

An array of data specifying for distances of the gene set expression profile by gene resampling, each row presents a expression profile of a specific gene label, each column represents a sample.

Usage

```
data(rand)
```

Format

These are both distances of gene expression profiles, each row represents a gene label, each column represents a sample

Source

They are derived using randdis

Examples

```
data(rand)
```

randdis	<i>Randomly generating some gene expression profiles by gene resampling</i>
---------	---

Description

A function which is used to generate the required number of gene expression profiles by permutation called gene resampling

Usage

```
randdis(z,minigenenum,randnum,setdis,normnum,Meth,resultname)
```


Arguments

z	An input matrix which contains the gene profiles
minigenenum	An integer indicating the number of gene to be permuted, the number of it has to be larger than the gene number that any gene function label has.
randnum	An integer represents the number of permutation
setdis	a character string indicating which method is to be used to compute the distances between case group and control group, avelinkdis or centdis is the choice to choose
normnum	an integer indecating the number of the case group
Meth	A character string indicates which method to be used to compute the distances between genes, euclidean and Manhattan is available to choose
resultname	character string which represents the name of the output file

Value

A text containing the distances only.

Author(s)

Li Jin, Huang Meilin

Examples

```
## Not run:
data(afExp)
##Randomly generating the gene expression profile and save it in the text named rand
randdis(afExp,500,10,avelinkdis,10,"euclidean","rand.txt")

## End(Not run)
```

valuep	<i>function which computes p values from different phenotypes by distance matrix</i>
--------	--

Description

this function is used to compute the p values from two matrixes which has been calculated before

Usage

```
valuep(MM,N,pname)
```

Arguments

MM	a matrix of gene phenotype distances which is generated before by the specific formula
N	a matrix of gene phenotype distances which is generated before randomly by gene resampling
pname	character string which represents the name of the output file

Value

A named matrix of p values and fdr p values.

Author(s)

Li Jin, Huang Meilin

Examples

```
## Not run:  
library(fdrtool)  
data(distance)  
data(rand)  
##compute the required p values  
valuep(distance,rand,"valuep.txt")  
  
## End(Not run)
```

Index

*Topic **datasets**

afExp, 3
b1, 4
distance, 6
rand, 8

*Topic **methods**

avelinkdis, 3
centdis, 4
distable, 5
functionall, 6
randdis, 8
valuep, 9

*Topic **package**

DBGSA-package, 2

afExp, 3
avelinkdis, 3

b1, 4

centdis, 4

DBGSA (DBGSA-package), 2
DBGSA-package, 2
distable, 5, 7
distance, 6

functionall, 6

rand, 8
randdis, 7, 8

valuep, 7, 9