

Package ‘DOSE’

December 11, 2024

Type Package

Title Disease Ontology Semantic and Enrichment analysis

Version 4.1.0

Maintainer Guangchuang Yu <guangchuangyu@gmail.com>

Description This package implements five methods proposed by Resnik, Schlicker, Jiang, Lin and Wang respectively for measuring semantic similarities among DO terms and gene products. Enrichment analyses including hypergeometric model and gene set enrichment analysis are also implemented for discovering disease associations of high-throughput biological data.

Depends R (>= 3.5.0)

Imports AnnotationDbi, BiocParallel, fgsea, ggplot2, GOSemSim (>= 2.31.2), methods, qvalue, reshape2, stats, utils, yulab.utils (>= 0.1.6)

Suggests prettydoc, clusterProfiler, gson (>= 0.0.5), knitr, memoise, org.Hs.eg.db, rmarkdown, testthat

VignetteBuilder knitr

ByteCompile true

License Artistic-2.0

Encoding UTF-8

URL <https://yulab-smu.top/contribution-knowledge-mining/>

BugReports <https://github.com/GuangchuangYu/DOSE/issues>

biocViews Annotation, Visualization, MultipleComparison, GeneSetEnrichment, Pathways, Software

RoxygenNote 7.3.2

git_url <https://git.bioconductor.org/packages/DOSE>

git_branch devel

git_last_commit 898262a

git_last_commit_date 2024-10-29

Repository Bioconductor 3.21

Date/Publication 2024-12-11

Author Guangchuang Yu [aut, cre],
 Li-Gen Wang [ctb],
 Vladislav Petyuk [ctb],
 Giovanni Dall'Olio [ctb]

Contents

DOSE-package	3
clusterSim	3
compareClusterResult-class	4
computeIC	5
DataSet	6
doseSim	6
enrichDGN	7
enrichDGNv	8
enrichDO	9
enricher_internal	10
enrichNCG	11
enrichResult-class	12
EXTID2NAME	13
gene2DO	13
geneID	14
geneInCategory	14
geneSim	15
gseaResult-class	16
GSEA_internal	17
gseDGN	18
gseDO	19
gseNCG	20
gsfilter	21
mclusterSim	21
parse_ratio	22
reexports	23
setReadable	23
show	24
simplot	24
summary	25
theme_dose	26

Index

27

DOSE-package

DOSE: Disease Ontology Semantic and Enrichment analysis

Description

This package implements five methods proposed by Resnik, Schlicker, Jiang, Lin and Wang respectively for measuring semantic similarities among DO terms and gene products. Enrichment analyses including hypergeometric model and gene set enrichment analysis are also implemented for discovering disease associations of high-throughput biological data.

Author(s)

Maintainer: Guangchuang Yu <guangchuangyu@gmail.com>

Other contributors:

- Li-Gen Wang <reeganwang020@gmail.com> [contributor]
- Vladislav Petyuk <petyuk@gmail.com> [contributor]
- Giovanni Dall'Olio <giovanni.dallolio@upf.edu> [contributor]

See Also

Useful links:

- <https://yulab-smu.top/biomedical-knowledge-mining-book/>
- Report bugs at <https://github.com/GuangchuangYu/DOSE/issues>

clusterSim

clusterSim

Description

semantic similarity between two gene clusters

Usage

```
clusterSim(  
  cluster1,  
  cluster2,  
  ont = "HDO",  
  organism = "hsa",  
  measure = "Wang",  
  combine = "BMA"  
)
```

Arguments

cluster1	a vector of gene IDs
cluster2	another vector of gene IDs
ont	one of "HDO", "HPO" and "MPO"
organism	one of "hsa" and "mmu"
measure	One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods.
combine	One of "max", "avg", "rcmax", "BMA" methods, for combining

Details

given two gene clusters, this function calculates semantic similarity between them.

Value

similarity

Author(s)

Yu Guangchuang

Examples

```
## Not run:  
cluster1 <- c("835", "5261", "241", "994")  
cluster2 <- c("307", "308", "317", "321", "506", "540", "378", "388", "396")  
clusterSim(cluster1, cluster2)  
  
## End(Not run)
```

compareClusterResult-class

Class "compareClusterResult" This class represents the comparison result of gene clusters by GO categories at specific level or GO enrichment analysis.

Description

Class "compareClusterResult" This class represents the comparison result of gene clusters by GO categories at specific level or GO enrichment analysis.

Slots

compareClusterResult cluster comparing result
geneClusters a list of genes
fun one of groupGO, enrichGO and enrichKEGG
gene2Symbol gene ID to Symbol
keytype Gene ID type
readable logical flag of gene ID in symbol or not.
.call function call
termsim Similarity between term
method method of calculating the similarity between nodes
dr dimension reduction result

Author(s)

Guangchuang Yu <https://yulab-smu.top>

See Also

[enrichResult](#)

computeIC *compute information content*

Description

compute information content

Usage

```
computeIC(ont = "HDO")
```

Arguments

ont one of "DO", "HPO" and "MPO"

Author(s)

Guangchuang Yu <https://yulab-smu.top>

 DataSet

Datasets

Description

Information content and DO term to entrez gene IDs mapping

doseSim

*doseSim***Description**

measuring similarities between two DO term vectors.

Usage

```
doseSim(DO1, DO2, measure = "Wang", ont = "HDO")
```

```
doSim(DO1, DO2, measure = "Wang", ont = "HDO")
```

Arguments

DO1	DO term, MPO term or HPO term vector
DO2	DO term, MPO term or HPO term vector
measure	one of "Wang", "Resnik", "Rel", "Jiang", "Lin", and "TCSS".
ont	one of "HDO", "HPO" and "MPO"

Details

provide two term vectors, this function will calculate their similarities.

Value

score matrix

Author(s)

Guangchuang Yu <https://yulab-smu.top>

enrichDGN *Enrichment analysis based on the DisGeNET* (<http://www.disgenet.org/>)

Description

given a vector of genes, this function will return the enrichment NCG categories with FDR control

Usage

```
enrichDGN(  
  gene,  
  pvalueCutoff = 0.05,  
  pAdjustMethod = "BH",  
  universe,  
  minGSSize = 10,  
  maxGSSize = 500,  
  qvalueCutoff = 0.2,  
  readable = FALSE  
)
```

Arguments

gene	a vector of entrez gene id
pvalueCutoff	pvalue cutoff
pAdjustMethod	one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"
universe	background genes
minGSSize	minimal size of genes annotated by NCG category for testing
maxGSSize	maximal size of each geneSet for analyzing
qvalueCutoff	qvalue cutoff
readable	whether mapping gene ID to gene Name

Value

A enrichResult instance

Author(s)

Guangchuang Yu

References

Janet et al. (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database* bav028 <http://database.oxfordjournals.org/content/2015/bav028.long>

`enrichDGNv`*enrichDGN*

Description

Enrichment analysis based on the DisGeNET (<http://www.disgenet.org/>)

Usage

```
enrichDGNv(  
  snp,  
  pvalueCutoff = 0.05,  
  pAdjustMethod = "BH",  
  universe,  
  minGSSize = 10,  
  maxGSSize = 500,  
  qvalueCutoff = 0.2,  
  readable = FALSE  
)
```

Arguments

<code>snp</code>	a vector of SNP
<code>pvalueCutoff</code>	pvalue cutoff
<code>pAdjustMethod</code>	one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"
<code>universe</code>	background genes
<code>minGSSize</code>	minimal size of genes annotated by NCG category for testing
<code>maxGSSize</code>	maximal size of each geneSet for analyzing
<code>qvalueCutoff</code>	qvalue cutoff
<code>readable</code>	whether mapping gene ID to gene Name

Details

given a vector of genes, this function will return the enrichment NCG categories with FDR control

Value

A `enrichResult` instance

Author(s)

Guangchuang Yu

References

Janet et al. (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database* bav028 <http://database.oxfordjournals.org/content/2015/bav028.long>

enrichDO

DO Enrichment Analysis

Description

Given a vector of genes, this function will return the enrichment DO categories with FDR control.

Usage

```
enrichDO(
  gene,
  ont = "HDO",
  organism = "hsa",
  pvalueCutoff = 0.05,
  pAdjustMethod = "BH",
  universe,
  minGSSize = 10,
  maxGSSize = 500,
  qvalueCutoff = 0.2,
  readable = FALSE
)
```

Arguments

gene	a vector of entrez gene id
ont	one of "HDO", "HPO" or "MPO".
organism	one of "hsa" and "mmu"
pvalueCutoff	pvalue cutoff
pAdjustMethod	one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"
universe	background genes
minGSSize	minimal size of genes annotated by NCG category for testing
maxGSSize	maximal size of each geneSet for analyzing
qvalueCutoff	qvalue cutoff
readable	whether mapping gene ID to gene Name

Value

A enrichResult instance.

Author(s)

Guangchuang Yu <https://yulab-smu.top>

See Also

[enrichResult-class](#)

Examples

```
data(geneList)
gene = names(geneList)[geneList > 1]
yy = enrichD0(gene, pvalueCutoff=0.05)
summary(yy)
```

enricher_internal *enrich.internal*

Description

internal method for enrichment analysis

Usage

```
enricher_internal(
  gene,
  pvalueCutoff,
  pAdjustMethod = "BH",
  universe = NULL,
  minGSSize = 10,
  maxGSSize = 500,
  qvalueCutoff = 0.2,
  USER_DATA
)
```

Arguments

gene	a vector of entrez gene id.
pvalueCutoff	Cutoff value of pvalue.
pAdjustMethod	one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"
universe	background genes, default is the intersection of the 'universe' with genes that have annotations. Users can set 'options(enrichment_force_universe = TRUE)' to force the 'universe' untouched.
minGSSize	minimal size of genes annotated by Ontology term for testing.
maxGSSize	maximal size of each geneSet for analyzing
qvalueCutoff	cutoff of qvalue
USER_DATA	ontology information

Details

using the hypergeometric model

Value

A `enrichResult` instance.

Author(s)

Guangchuang Yu <https://yulab-smu.top>

enrichNCG

enrichNCG

Description

Enrichment analysis based on the Network of Cancer Genes database (<http://ncg.kcl.ac.uk/>)

Usage

```
enrichNCG(
  gene,
  pvalueCutoff = 0.05,
  pAdjustMethod = "BH",
  universe,
  minGSSize = 10,
  maxGSSize = 500,
  qvalueCutoff = 0.2,
  readable = FALSE
)
```

Arguments

<code>gene</code>	a vector of entrez gene id
<code>pvalueCutoff</code>	pvalue cutoff
<code>pAdjustMethod</code>	one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"
<code>universe</code>	background genes
<code>minGSSize</code>	minimal size of genes annotated by NCG category for testing
<code>maxGSSize</code>	maximal size of each <code>geneSet</code> for analyzing
<code>qvalueCutoff</code>	qvalue cutoff
<code>readable</code>	whether mapping gene ID to gene Name

Details

given a vector of genes, this function will return the enrichment NCG categories with FDR control

Value

A enrichResult instance

Author(s)

Guangchuang Yu

enrichResult-class	<i>Class "enrichResult" This class represents the result of enrichment analysis.</i>
--------------------	--

Description

Class "enrichResult" This class represents the result of enrichment analysis.

Slots

result enrichment analysis
 pvalueCutoff pvalueCutoff
 pAdjustMethod pvalue adjust method
 qvalueCutoff qvalueCutoff
 organism only "human" supported
 ontology biological ontology
 gene Gene IDs
 keytype Gene ID type
 universe background gene
 gene2Symbol mapping gene to Symbol
 geneSets gene sets
 readable logical flag of gene ID in symbol or not.
 termsim Similarity between term
 method method of calculating the similarity between nodes
 dr dimension reduction result

Author(s)

Guangchuang Yu <https://yulab-smu.top>

See Also

[enrichDO](#)

EXTID2NAME	<i>EXTID2NAME</i>
------------	-------------------

Description

mapping gene ID to gene Symbol

Usage

```
EXTID2NAME(OrgDb, geneID, keytype)
```

Arguments

OrgDb	OrgDb
geneID	entrez gene ID
keytype	keytype

Value

gene symbol

Author(s)

Guangchuang Yu <https://yulab-smu.top>

gene2DO	<i>convert Gene ID to DO Terms</i>
---------	------------------------------------

Description

provide gene ID, this function will convert to the corresponding DO Terms

Usage

```
gene2DO(gene, organism = "hsa", ont = "HDO")
```

Arguments

gene	entrez gene ID
organism	organism
ont	ont

Value

DO Terms

Author(s)

Guangchuang Yu <https://yulab-smu.top>

geneID	<i>geneID generic</i>
--------	-----------------------

Description

geneID generic

Usage

```
geneID(x)
```

Arguments

x enrichResult object

Value

'geneID' return the 'geneID' column of the enriched result which can be converted to data.frame via 'as.data.frame'

Examples

```
data(geneList, package="DOSE")
de <- names(geneList)[1:100]
x <- enrichDO(de)
geneID(x)
```

geneInCategory	<i>geneInCategory generic</i>
----------------	-------------------------------

Description

geneInCategory generic

Usage

```
geneInCategory(x)
```

Arguments

x enrichResult

Value

'geneInCategory' return a list of genes, by splitting the input gene vector to enriched functional categories

Examples

```
data(geneList, package="DOSE")
de <- names(geneList)[1:100]
x <- enrichDO(de)
geneInCategory(x)
```

 geneSim

geneSim

Description

measuring similarities between two gene vectors.

Usage

```
geneSim(
  geneID1,
  geneID2 = NULL,
  ont = "HDO",
  organism = "hsa",
  measure = "Wang",
  combine = "BMA"
)
```

Arguments

geneID1	entrez gene vector
geneID2	entrez gene vector
ont	one of "HDO" and "MPO"
organism	one of "hsa" and "mmu"
measure	one of "Wang", "Resnik", "Rel", "Jiang", and "Lin".
combine	One of "max", "avg", "rcmax", "BMA" methods, for combining semantic similarity scores of multiple DO terms associated with gene/protein.

Details

provide two entrez gene vectors, this function will calculate their similarity.

Value

score matrix

Author(s)

Guangchuang Yu <https://yulab-smu.top>

Examples

```
g <- c("835", "5261", "241", "994")
geneSim(g)
```

gseaResult-class

Class "gseaResult" This class represents the result of GSEA analysis

Description

Class "gseaResult" This class represents the result of GSEA analysis

Slots

result GSEA analysis

organism organism

setType setType

geneSets geneSets

geneList order rank geneList

keytype ID type of gene

permScores permutation scores

params parameters

gene2Symbol gene ID to Symbol

readable whether convert gene ID to symbol

dr dimension reduction result

Author(s)

Guangchuang Yu <https://yulab-smu.top>

GSEA_internal *GSEA_internal*

Description

generic function for gene set enrichment analysis

Usage

```
GSEA_internal(
  geneList,
  exponent,
  minGSSize,
  maxGSSize,
  eps,
  pvalueCutoff,
  pAdjustMethod,
  verbose,
  seed = FALSE,
  USER_DATA,
  by = "fgsea",
  ...
)
```

Arguments

geneList	order ranked geneList
exponent	weight of each step
minGSSize	minimal size of each geneSet for analyzing
maxGSSize	maximal size of each geneSet for analyzing
eps	This parameter sets the boundary for calculating the p value.
pvalueCutoff	p value Cutoff
pAdjustMethod	p value adjustment method
verbose	print message or not
seed	set seed inside the function to make result reproducible. FALSE by default.
USER_DATA	annotation data
by	one of 'fgsea' or 'DOSE'
...	other parameter

Value

gseaResult object

Author(s)

Yu Guangchuang

gseDGN

DisGeNET Gene Set Enrichment Analysis

Description

perform gsea analysis

Usage

```
gseDGN(  
  geneList,  
  exponent = 1,  
  minGSSize = 10,  
  maxGSSize = 500,  
  pvalueCutoff = 0.05,  
  pAdjustMethod = "BH",  
  verbose = TRUE,  
  seed = FALSE,  
  by = "fgsea",  
  ...  
)
```

Arguments

geneList	order ranked geneList
exponent	weight of each step
minGSSize	minimal size of each geneSet for analyzing
maxGSSize	maximal size of each geneSet for analyzing
pvalueCutoff	pvalue Cutoff
pAdjustMethod	p value adjustment method
verbose	print message or not
seed	logical
by	one of 'fgsea' or 'DOSE'
...	other parameter

Value

gseaResult object

Author(s)

Yu Guangchuang

gseDO

DO Gene Set Enrichment Analysis

Description

perform gsea analysis

Usage

```
gseDO(  
  geneList,  
  ont = "HDO",  
  organism = "hsa",  
  exponent = 1,  
  minGSSize = 10,  
  maxGSSize = 500,  
  pvalueCutoff = 0.05,  
  pAdjustMethod = "BH",  
  verbose = TRUE,  
  seed = FALSE,  
  by = "fgsea",  
  ...  
)
```

Arguments

geneList	order ranked geneList
ont	one of "HDO", "HPO" or "MPO"
organism	one of "hsa" and "mmu"
exponent	weight of each step
minGSSize	minimal size of each geneSet for analyzing
maxGSSize	maximal size of each geneSet for analyzing
pvalueCutoff	pvalue Cutoff
pAdjustMethod	p value adjustment method
verbose	print message or not
seed	logical
by	one of 'fgsea' or 'DOSE'
...	other parameter

Value

gseaResult object

Author(s)

Yu Guangchuang

gseNCG

NCG Gene Set Enrichment Analysis

Description

perform gsea analysis

Usage

```
gseNCG(  
  geneList,  
  exponent = 1,  
  minGSSize = 10,  
  maxGSSize = 500,  
  pvalueCutoff = 0.05,  
  pAdjustMethod = "BH",  
  verbose = TRUE,  
  seed = FALSE,  
  by = "fgsea",  
  ...  
)
```

Arguments

geneList	order ranked geneList
exponent	weight of each step
minGSSize	minimal size of each geneSet for analyzing
maxGSSize	maximal size of each geneSet for analyzing
pvalueCutoff	pvalue Cutoff
pAdjustMethod	p value adjustment method
verbose	print message or not
seed	logical
by	one of 'fgsea' or 'DOSE'
...	other parameter

Value

gseaResult object

Author(s)

Yu Guangchuang

gsfilter	<i>gsfilter</i>
----------	-----------------

Description

filter enriched result by gene set size or gene count

Usage

```
gsfilter(x, by = "GSSize", min = NA, max = NA)
```

Arguments

x	instance of enrichResult or compareClusterResult
by	one of 'GSSize' or 'Count'
min	minimal size
max	maximal size

Value

update object

Author(s)

Guangchuang Yu

mclusterSim	<i>mclusterSim</i>
-------------	--------------------

Description

Pairwise semantic similarity for a list of gene clusters

Usage

```
mclusterSim(  
  clusters,  
  ont = "HDO",  
  organism = "hsa",  
  measure = "Wang",  
  combine = "BMA"  
)
```

Arguments

clusters	A list of gene clusters
ont	one of "HDO", "HPO" and "MPO"
organism	organism
measure	one of "Wang", "Resnik", "Rel", "Jiang", and "Lin".
combine	One of "max", "avg", "rcmax", "BMA" methods, for combining semantic similarity scores of multiple DO terms associated with gene/protein.

Value

similarity matrix

Author(s)

Guangchuang Yu

Examples

```
## Not run:
cluster1 <- c("835", "5261", "241")
cluster2 <- c("578", "582")
cluster3 <- c("307", "308", "317")
clusters <- list(a=cluster1, b=cluster2, c=cluster3)
mclusterSim(clusters, measure="Wang")

## End(Not run)
```

parse_ratio

parse_ratio

Description

parse character ratio to double value, such as 1/5 to 0.2

Usage

```
parse_ratio(ratio)
```

Arguments

ratio	character vector of ratio to parse
-------	------------------------------------

Value

A numeric vector (double) of parsed ratio

Author(s)

Guangchuang Yu

reexports	<i>Objects exported from other packages</i>
-----------	---

Description

These objects are imported from other packages. Follow the links below to see their documentation.

ggplot2 [facet_grid](#)

setReadable	<i>setReadable</i>
-------------	--------------------

Description

mapping geneID to gene Symbol

Usage

```
setReadable(x, OrgDb, keyType = "auto")
```

Arguments

x	enrichResult Object
OrgDb	OrgDb
keyType	keyType of gene

Value

enrichResult Object

Author(s)

Yu Guangchuang

show *show method*

Description

show method for gseaResult instance
show method for enrichResult instance

Usage

```
show(object)
```

```
show(object)
```

Arguments

object A enrichResult instance.

Value

message

message

Author(s)

Guangchuang Yu <https://yulab-smu.top>

simplot *simplot*

Description

plotting similarity matrix

Usage

```
simplot(  
  sim,  
  xlab = "",  
  ylab = "",  
  color.low = "white",  
  color.high = "red",  
  labs = TRUE,  
  digits = 2,  
  labs.size = 3,  
  font.size = 14  
)
```


Arguments

sim	similarity matrix
xlab	xlab
ylab	ylab
color.low	color of low value
color.high	color of high value
labs	logical, add text label or not
digits	round digit numbers
labs.size	lable size
font.size	font size

Value

ggplot object

Author(s)

Yu Guangchuang

summary	<i>summary method</i>
---------	-----------------------

Description

summary method for gseaResult instance
summary method for enrichResult instance

Usage

```
summary(object, ...)
```

```
summary(object, ...)
```

Arguments

object	A enrichResult instance.
...	additional parameter

Value

A data frame
A data frame

Author(s)

Guangchuang Yu <https://guangchuangyu.github.io>

Guangchuang Yu <http://guangchuangyu.github.io>

theme_dose

theme_dose

Description

ggplot theme of DOSE

Usage

```
theme_dose(font.size = 14)
```

Arguments

font.size font size

Value

ggplot theme

Examples

```
library(ggplot2)
qplot(1:10) + theme_dose()
```

Index

- * **classes**
 - compareClusterResult-class, 4
 - enrichResult-class, 12
 - gseaResult-class, 16
- * **datasets**
 - DataSet, 6
- * **internal**
 - DOSE-package, 3
 - reexports, 23
- * **manip**
 - enrichDO, 9
 - enricher_internal, 10
 - gseDGN, 18
 - gseDO, 19
 - gseNCG, 20
- clusterSim, 3
- compareClusterResult-class, 4
- computeIC, 5
- DataSet, 6
- DGN_EXTID2PATHID (DataSet), 6
- DGN_PATHID2EXTID (DataSet), 6
- DGN_PATHID2NAME (DataSet), 6
- DOSE (DOSE-package), 3
- DOSE-package, 3
- doseSim, 6
- doSim (doseSim), 6
- enrichDGN, 7
- enrichDGNv, 8
- enrichDO, 9, 12
- enricher_internal, 10
- enrichNCG, 11
- enrichResult, 5
- enrichResult-class, 12
- EXTID2NAME, 13
- facet_grid, 23
- facet_grid (reexports), 23
- gene2DO, 13
- geneID, 14
- geneInCategory, 14
- geneList (DataSet), 6
- geneSim, 15
- GSEA_internal, 17
- gseahResult-class (gseaResult-class), 16
- gseaResult-class, 16
- gseDGN, 18
- gseDO, 19
- gseNCG, 20
- gsfilter, 21
- mclusterSim, 21
- NCG_EXTID2PATHID (DataSet), 6
- NCG_PATHID2EXTID (DataSet), 6
- NCG_PATHID2NAME (DataSet), 6
- parse_ratio, 22
- plot, compareClusterResult-method (compareClusterResult-class), 4
- reexports, 23
- setReadable, 23
- show, 24
- show, compareClusterResult-method (compareClusterResult-class), 4
- show, enrichResult-method (enrichResult-class), 12
- show, gseaResult-method (gseaResult-class), 16
- simplot, 24
- summary, 25
- summary, compareClusterResult-method (compareClusterResult-class), 4
- summary, enrichResult-method (enrichResult-class), 12
- summary, gseaResult-method (gseaResult-class), 16

theme_dose, [26](#)

VDGN_EXTID2PATHID (DataSet), [6](#)

VDGN_PATHID2EXTID (DataSet), [6](#)

VDGN_PATHID2NAME (DataSet), [6](#)