

Package ‘DOSE’

January 23, 2026

Type Package

Title Disease Ontology Semantic and Enrichment analysis

Version 4.5.1

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, enrichit (>= 0.0.4), ggplot2, GOSemSim (>= 2.37.0), methods, reshape2, utils, yulab.utils (> 0.2.2)

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 7.3.3

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

git_branch devel

git_last_commit 3414fdc

git_last_commit_date 2025-12-22

Repository Bioconductor 3.23
Date/Publication 2026-01-23
Author Guangchuang Yu [aut, cre],
Li-Gen Wang [ctb],
Vladislav Petyuk [ctb],
Giovanni Dall'Olio [ctb]

Contents

DOSE-package	2
clusterSim	3
computeIC	4
DataSet	4
doseSim	5
dose_params	5
enrichDGN	6
enrichDGNv	7
enrichDO	8
enrichNCG	9
gene2DO	10
geneSim	11
gseDGN	12
gseDO	13
gseNCG	14
mclusterSim	15
reexports	16
simplot	17
theme_dose	18
Index	19

DOSE-package	<i>DOSE: Disease Ontology Semantic and Enrichment analysis</i>
--------------	--

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/contribution-knowledge-mining/>
- 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)
```

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	<i>doseSim</i>
---------	----------------

Description

measuring similarities between two DO term vectors.

Usage

```
doseSim(DOID1, DOID2, measure = "Wang", ont = "HDO")  
  
doSim(DOID1, DOID2, measure = "Wang", ont = "HDO")
```

Arguments

DOID1	DO term, MPO term or HPO term vector
DOID2	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>

dose_params	<i>Shared parameters for DOSE functions</i>
-------------	---

Description

Shared parameters for DOSE functions

Arguments

gene	a vector of entrez gene id
organism	one of "hsa" and "mmu"
ont	one of "HDO", "HPO" or "MPO"
pvalueCutoff	pvalue cutoff
pAdjustMethod	one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"
universe	background genes
minGSSize	minimal size of genes annotated by ontology term for testing
maxGSSize	maximal size of each geneSet for analyzing
qvalueCutoff	qvalue cutoff
readable	whether mapping gene ID to gene Name
geneList	order ranked geneList
exponent	weight of each step
nPerm	permutation numbers
verbose	print message or not
adaptive	logical, use adaptive permutation or not (default: FALSE)
minPerm	minimum number of permutations for adaptive mode (default: 1000)
maxPerm	maximum number of permutations for adaptive mode (default: 10000)
method	method of GSEA, one of "multilevel", "permute", "sample"

enrichDGN	<i>Enrichment analysis based on the DisGeNET</i> (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 ontology term 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

snp	a vector of SNP
pvalueCutoff	pvalue cutoff
pAdjustMethod	one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"
universe	background genes
minGSSize	minimal size of genes annotated by ontology term for testing
maxGSSize	maximal size of each geneSet for analyzing
qvalueCutoff	qvalue cutoff
readable	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 ontology term 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>

Examples

```

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

```

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

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 ontology term for testing
maxGSSize	maximal size of each geneSet for analyzing
qvalueCutoff	qvalue cutoff
readable	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

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>

geneSim

geneSim

Description

measuring similarities bewteen 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)
```

gseDGN

DisGeNET Gene Set Enrichment Analysis

Description

perform gsea analysis

Usage

```
gseDGN(
  geneList,
  exponent = 1,
  nPerm = 1000,
  minGSSize = 10,
  maxGSSize = 500,
  pvalueCutoff = 0.05,
  pAdjustMethod = "BH",
  verbose = TRUE,
  method = "multilevel",
  adaptive = FALSE,
  minPerm = 1000,
  maxPerm = 10000,
  ...
)
```

Arguments

geneList	order ranked geneList
exponent	weight of each step
nPerm	permutation numbers
minGSSize	minimal size of genes annotated by ontology term for testing
maxGSSize	maximal size of each geneSet for analyzing
pvalueCutoff	pvalue cutoff

pAdjustMethod	one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"
verbose	print message or not
method	method of GSEA, one of "multilevel", "permute", "sample"
adaptive	logical, use adaptive permutation or not (default: FALSE)
minPerm	minimum number of permutations for adaptive mode (default: 1000)
maxPerm	maximum number of permutations for adaptive mode (default: 10000)
...	other parameter

Value

gseaResult object

Author(s)

Guangchuang Yu

gseDO

DO Gene Set Enrichment Analysis

Description

perform gsea analysis

Usage

```
gseDO(
  geneList,
  ont = "HDO",
  organism = "hsa",
  exponent = 1,
  nPerm = 1000,
  minGSSize = 10,
  maxGSSize = 500,
  pvalueCutoff = 0.05,
  pAdjustMethod = "BH",
  verbose = TRUE,
  method = "multilevel",
  adaptive = FALSE,
  minPerm = 1000,
  maxPerm = 10000,
  ...
)
```

Arguments

geneList	order ranked geneList
ont	one of "HDO", "HPO" or "MPO"
organism	one of "hsa" and "mmu"
exponent	weight of each step
nPerm	permutation numbers
minGSSize	minimal size of genes annotated by ontology term for testing
maxGSSize	maximal size of each geneSet for analyzing
pvalueCutoff	pvalue cutoff
pAdjustMethod	one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"
verbose	print message or not
method	method of GSEA, one of "multilevel", "permute", "sample"
adaptive	logical, use adaptive permutation or not (default: FALSE)
minPerm	minimum number of permutations for adaptive mode (default: 1000)
maxPerm	maximum number of permutations for adaptive mode (default: 10000)
...	other parameter

Value

gseaResult object

Author(s)

Guangchuang Yu

gseNCG

NCG Gene Set Enrichment Analysis

Description

perform gsea analysis

Usage

```
gseNCG(
  geneList,
  exponent = 1,
  nPerm = 1000,
  minGSSize = 10,
  maxGSSize = 500,
  pvalueCutoff = 0.05,
  pAdjustMethod = "BH",
```

```
    verbose = TRUE,  
    method = "multilevel",  
    adaptive = FALSE,  
    minPerm = 1000,  
    maxPerm = 10000,  
    ...  
  )
```

Arguments

geneList	order ranked geneList
exponent	weight of each step
nPerm	permutation numbers
minGSSize	minimal size of genes annotated by ontology term for testing
maxGSSize	maximal size of each geneSet for analyzing
pvalueCutoff	pvalue cutoff
pAdjustMethod	one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"
verbose	print message or not
method	method of GSEA, one of "multilevel", "permute", "sample"
adaptive	logical, use adaptive permutation or not (default: FALSE)
minPerm	minimum number of permutations for adaptive mode (default: 1000)
maxPerm	maximum number of permutations for adaptive mode (default: 10000)
...	other parameter

Value

gseaResult 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

<code>clusters</code>	A list of gene clusters
<code>ont</code>	one of "HDO", "HPO" and "MPO"
<code>organism</code>	organism
<code>measure</code>	one of "Wang", "Resnik", "Rel", "Jiang", and "Lin".
<code>combine</code>	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)
```

reexports

Objects exported from other packages

Description

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

ggplot2 [facet_grid](#)

`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

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

Value

ggplot object

Author(s)

Yu Guangchuang

theme_dose	<i>theme_dose</i>
------------	-------------------

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

- * **datasets**
 - DataSet, [4](#)
- * **internal**
 - DOSE-package, [2](#)
 - reexports, [16](#)
- * **manip**
 - enrichDO, [8](#)
 - gseDGN, [12](#)
 - gseDO, [13](#)
 - gseNCG, [14](#)
- clusterSim, [3](#)
- computeIC, [4](#)
- DataSet, [4](#)
- DGN_EXTID2PATHID (DataSet), [4](#)
- DGN_PATHID2EXTID (DataSet), [4](#)
- DGN_PATHID2NAME (DataSet), [4](#)
- DOSE (DOSE-package), [2](#)
- DOSE-package, [2](#)
- dose_params, [5](#)
- doseSim, [5](#)
- doSim (doseSim), [5](#)
- enrichDGN, [6](#)
- enrichDGNv, [7](#)
- enrichDO, [8](#)
- enrichNCG, [9](#)
- facet_grid, [16](#)
- facet_grid (reexports), [16](#)
- gene2DO, [10](#)
- geneList (DataSet), [4](#)
- geneSim, [11](#)
- gseDGN, [12](#)
- gseDO, [13](#)
- gseNCG, [14](#)
- mclusterSim, [15](#)
- NCG_EXTID2PATHID (DataSet), [4](#)
- NCG_PATHID2EXTID (DataSet), [4](#)
- NCG_PATHID2NAME (DataSet), [4](#)
- reexports, [16](#)
- simplot, [17](#)
- theme_dose, [18](#)
- VDGN_EXTID2PATHID (DataSet), [4](#)
- VDGN_PATHID2EXTID (DataSet), [4](#)
- VDGN_PATHID2NAME (DataSet), [4](#)