

# Package ‘DOSE’

April 15, 2020

**Type** Package

**Title** Disease Ontology Semantic and Enrichment analysis

**Version** 3.12.0

**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 (>= 3.4.0)

**Imports** AnnotationDbi, BiocParallel, DO.db, fgsea, ggplot2, GOSemSim (>= 2.0.0), methods, qvalue, reshape2, S4Vectors, stats, utils

**Suggests** prettydoc, clusterProfiler, knitr, org.Hs.eg.db, testthat

**VignetteBuilder** knitr

**ByteCompile** true

**License** Artistic-2.0

**URL** <https://guangchuangyu.github.io/software/DOSE>

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

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

**RoxygenNote** 6.1.1

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

**git\_branch** RELEASE\_3\_10

**git\_last\_commit** 94422dc

**git\_last\_commit\_date** 2019-10-29

**Date/Publication** 2020-04-14

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

**R topics documented:**

DOSE-package	2
clusterSim	3
computeIC	4
DataSet	4
doSim	5
enrichDGN	5
enrichDGNv	6
enrichDO	7
enricher_internal	8
enrichNCG	9
enrichResult-class	9
EXTID2NAME	10
fortify.enrichResult	11
gene2DO	11
geneID	12
geneInCategory	12
geneSim	13
gseaResult-class	13
GSEA_internal	14
gseDGN	15
gseDO	15
gseNCG	16
gsfilter	17
mclusterSim	17
parse_ratio	18
rebuildAnnoData	19
setReadable	19
show	20
simplot	20
summary	21
theme_dose	22

**Index** **23**


---

DOSE-package	<i>Disease Ontology Semantic and Enrichment analysis Implemented five methods proposed by Resnik, Schlicker, Jiang, Lin and Wang respectively for measuring DO semantic similarities, and hypergeometric test for enrichment analysis.</i>
--------------	--

---

**Description**

This package is designed to estimate DO-based semantic similarity measurement and enrichment analysis.

**Details**

Package: DOSE  
Type: Package  
Version: 2.3.5  
Date: 2-27-2012  
biocViews: Bioinformatics, Annotation  
Depends:  
Imports: methods, AnnotationDbi, DO.db  
Suggests: clusterProfiler, GOSemSim  
License: Artistic-2.0

**Author(s)**

Guangchuang Yu, Li-Gen Wang

Maintainer: Guangchuang Yu <guangchuangyu@gmail.com>

**See Also**

[enrichResult](#)

---

clusterSim

*clusterSim*

---

**Description**

semantic similarity between two gene clusters

**Usage**

```
clusterSim(cluster1, cluster2, measure = "Wang", combine = "BMA")
```

**Arguments**

cluster1	a vector of gene IDs
cluster2	another vector of gene IDs
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**

```
cluster1 <- c("835", "5261", "241", "994")
cluster2 <- c("307", "308", "317", "321", "506", "540", "378", "388", "396")
clusterSim(cluster1, cluster2)
```

---

computeIC	<i>compute information content</i>
-----------	------------------------------------

---

**Description**

compute information content

**Usage**

```
computeIC(ont = "DO", organism = "human")
```

**Arguments**

ont	"DO"
organism	"human"

**Author(s)**

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

---

DataSet	<i>Datasets</i>
---------	-----------------

---

**Description**

Information content and DO term to entrez gene IDs mapping

---

doSim	<i>doSim</i>
-------	--------------

---

**Description**

measuring similarities between two DO term vectors.

**Usage**

```
doSim(DO1D1, DO1D2, measure = "Wang")
```

**Arguments**

DO1D1	DO term vector
DO1D2	DO term vector
measure	one of "Wang", "Resnik", "Rel", "Jiang", and "Lin".

**Details**

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

**Value**

score matrix

**Author(s)**

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

---

enrichDGN	<i>Enrichment analysis based on the DisGeNET</i> ( <a href="http://www.disgenet.org/">http://www.disgenet.org/</a> )
-----------	--

---

**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**

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 NCG category 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 = "DO", 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 DO or DOLite.
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 <http://guangchuangyu.github.io>

**See Also**[enrichResult-class](#)**Examples**

```
data(geneList)
gene = names(geneList)[geneList > 1]
yy = enrichDO(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
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 <http://guangchuangyu.github.io>



---

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 NCG category 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

---

enrichResult-class

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


---

**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.

**Author(s)**

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

**See Also**

[enrichDO](#)

---

EXTID2NAME

*EXTID2NAME*

---

**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 <http://guangchuangyu.github.io>

---

 fortify.enrichResult *fortify*


---

**Description**

convert enrichResult object for ggplot2

**Usage**

```
## S3 method for class 'enrichResult'
fortify(model, data, showCategory = 5,
  by = "Count", order = FALSE, drop = FALSE, split = NULL, ...)
```

**Arguments**

model	enrichResult object
data	not use here
showCategory	Category numbers to show
by	one of Count and GeneRatio
order	logical
drop	logical
split	separate result by 'split' variable
...	additional parameter

---

 gene2DO *convert Gene ID to DO Terms*


---

**Description**

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

**Usage**

```
gene2DO(gene)
```

**Arguments**

gene	entrez gene ID
------	----------------

**Value**

DO Terms

**Author(s)**

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

---

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	<i>geneSim</i>
---------	----------------

---

**Description**

measuring similarities between two gene vectors.

**Usage**

```
geneSim(geneID1, geneID2 = NULL, measure = "Wang", combine = "BMA")
```

**Arguments**

geneID1	entrez gene vector
geneID2	entrez gene vector
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 <http://ygc.name>

---

gseaResult-class	<i>Class "gseaResult" This class represents the result of GSEA analysis</i>
------------------	---

---

**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

**Author(s)**

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

---

GSEA\_internal

*GSEA\_internal*

---

**Description**

generic function for gene set enrichment analysis

**Usage**

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

**Arguments**

geneList	order ranked geneList
exponent	weight of each step
nPerm	permutation numbers
minGSSize	minimal size of each geneSet for analyzing
maxGSSize	maximal size of each geneSet for analyzing
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'

**Value**

gseaResult object

**Author(s)**

Yu Guangchuang

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, seed = FALSE, by = "fgsea")
```

**Arguments**

geneList	order ranked geneList
exponent	weight of each step
nPerm	permutation numbers
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'

**Value**

gseaResult object

**Author(s)**

Yu Guangchuang

gseDO

*DO Gene Set Enrichment Analysis***Description**

perform gsea analysis

**Usage**

```
gseDO(geneList, exponent = 1, nPerm = 1000, 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
nPerm	permutation numbers
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'

**Value**

gseaResult object

**Author(s)**

Yu Guangchuang

---

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, seed = FALSE, by = "fgsea")
```

**Arguments**

geneList	order ranked geneList
exponent	weight of each step
nPerm	permutation numbers
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'



**Value**

gseaResult object

**Author(s)**

Yu Guangchuang

---

*gsfilter**gsfilter*

---

**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**mclusterSim*

---

**Description**

Pairwise semantic similarity for a list of gene clusters

**Usage**

mclusterSim(clusters, measure = "Wang", combine = "BMA")

**Arguments**

clusters	A list of gene clusters
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)**

Yu Guangchuang

**Examples**

```
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")
```

---

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

---

rebuildAnnoData	<i>rebuiding annotation data</i>
-----------------	----------------------------------

---

**Description**

rebuilding entrez and DO mapping datasets

**Usage**

```
rebuildAnnoData(file)
```

**Arguments**

file	do_rif.human.txt
------	------------------

**Author(s)**

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

---

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	<i>show method</i>
------	--------------------

---

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

---

simplot	<i>simplot</i>
---------	----------------

---

**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

*summary method*

---

**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	<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

- \*Topic **classes**
  - enrichResult-class, [9](#)
  - gseaResult-class, [13](#)
- \*Topic **datasets**
  - DataSet, [4](#)
- \*Topic **manip**
  - enrichDO, [7](#)
  - enricher\_internal, [8](#)
  - gseDGN, [15](#)
  - gseDO, [15](#)
  - gseNCG, [16](#)
- \*Topic **package**
  - DOSE-package, [2](#)
  
- clusterSim, [3](#)
- computeIC, [4](#)
  
- DataSet, [4](#)
- DGN\_EXTID2PATHID (DataSet), [4](#)
- DGN\_PATHID2EXTID (DataSet), [4](#)
- DGN\_PATHID2NAME (DataSet), [4](#)
- DO2ALLEG (DataSet), [4](#)
- DO2EG (DataSet), [4](#)
- DOIC (DataSet), [4](#)
- DOLite2EG (DataSet), [4](#)
- DOLiteTerm (DataSet), [4](#)
- DOSE (DOSE-package), [2](#)
- DOSE-package, [2](#)
- doSim, [5](#)
- dotbl (DataSet), [4](#)
  
- EG2ALLDO (DataSet), [4](#)
- EG2DO (DataSet), [4](#)
- EG2DOLite (DataSet), [4](#)
- enrichDGN, [5](#)
- enrichDGNv, [6](#)
- enrichDO, [7](#), [10](#)
- enricher\_internal, [8](#)
- enrichNCG, [9](#)
- enrichResult, [3](#)
- enrichResult-class, [9](#)
- EXTID2NAME, [10](#)
  
- fortify.enrichResult, [11](#)
  
- gene2DO, [11](#)
- geneID, [12](#)
- geneInCategory, [12](#)
- geneList (DataSet), [4](#)
- geneSim, [13](#)
- GSEA\_internal, [14](#)
- gseahResult-class (gseaResult-class), [13](#)
- gseaResult-class, [13](#)
- gseDGN, [15](#)
- gseDO, [15](#)
- gseNCG, [16](#)
- gsfilter, [17](#)
  
- mclusterSim, [17](#)
  
- NCG\_EXTID2PATHID (DataSet), [4](#)
- NCG\_PATHID2EXTID (DataSet), [4](#)
- NCG\_PATHID2NAME (DataSet), [4](#)
  
- parse\_ratio, [18](#)
  
- rebuildAnnoData, [19](#)
  
- setReadable, [19](#)
- show, [20](#)
- show, enrichResult-method (enrichResult-class), [9](#)
- show, gseaResult-method (gseaResult-class), [13](#)
- simplot, [20](#)
- summary, [21](#)
- summary, enrichResult-method (enrichResult-class), [9](#)
- summary, gseaResult-method (gseaResult-class), [13](#)
  
- theme\_dose, [22](#)
  
- VDGN\_EXTID2PATHID (DataSet), [4](#)
- VDGN\_PATHID2EXTID (DataSet), [4](#)
- VDGN\_PATHID2NAME (DataSet), [4](#)