Package 'DOSE'

November 12, 2026

Type Package

Title Disease Ontology Semantic and Enrichment analysis

Version 4.7.4

Maintainer Guangchuang Yu <guangchuangyu@gmail.com>

Description Provides ontology-aware methods for disease and phenotype knowledge mining. DOSE supports semantic similarity analysis of disease and phenotype ontology terms, genes, and gene clusters using methods including Resnik, Schlicker, Jiang, Lin, and Wang. It also provides over-representation analysis and gene set enrichment analysis for interpreting gene vectors and ranked gene lists in disease, phenotype, and cancer contexts.

Depends R (>= 3.5.0)

Imports AnnotationDbi, enrichit (>= 0.2.2), ggplot2, GOSemSim (>= 2.39.2), gson (>= 0.0.5), memoise, methods, reshape2, utils, yulab.utils (> 0.2.2)

Suggests aisdk, clusterProfiler, knitr, org.Hs.eg.db, org.Mm.eg.db, quarto, testthat

VignetteBuilder quarto

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

Config/roxygen2/version 8.1.0

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

git_branch devel

git_last_commit c90a0ae

git_last_commit_date 2026-11-11

Repository Bioconductor 3.24

Date/Publication 2026-11-12

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

Contents

DOSE-package . . . 2
as.data.frame,doseInterpretResult-method . . . 3
clusterSim . . . 4
computeIC . . . 5
DataSet . . . 5
doseInterpretResult . . . 5
doseInterpretResult-class . . . 6
doseSim . . . 6
dose_params . . . 7
enrichDO . . . 8
enrichNCG . . . 9
evidence . . . 10
explainDisease . . . 10
gene2DO . . . 11
geneSim . . . 12
gseDO . . . 13
gseNCG . . . 14
interpretDisease . . . 15
mclusterSim . . . 16
rankMouseModels . . . 17
reexports . . . 18
show,doseInterpretResult-method . . . 18
simplot . . . 18
summary,doseInterpretResult-method . . . 19
theme_dose . . . 20

Index **21**

`DOSE-package` *DOSE: Disease Ontology Semantic and Enrichment analysis*

Description

Provides ontology-aware methods for disease and phenotype knowledge mining. DOSE supports semantic similarity analysis of disease and phenotype ontology terms, genes, and gene clusters using methods including Resnik, Schlicker, Jiang, Lin, and Wang. It also provides over-representation analysis and gene set enrichment analysis for interpreting gene vectors and ranked gene lists in disease, phenotype, and cancer contexts.

Author(s)

Maintainer: Guangchuang Yu <guangchuangyu@gmail.com>

Authors:

- 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

---

as.data.frame,doseInterpretResult-method

Coerce a canonical interpretation result to a data.frame

Description

Coerce a canonical interpretation result to a data.frame

Usage

```
## S4 method for signature 'doseInterpretResult'
as.data.frame(x, row.names = NULL, optional = FALSE, ...)
```

Arguments

x	A 'doseInterpretResult' object.
row.names, optional, ...	Passed to [base::as.data.frame()].

Value

A data.frame of ranked result rows.

clusterSim *clusterSim*

Description

semantic similarity between two gene clusters

Usage

```
clusterSim(
  cluster1,
  cluster2,
  ont = "HDO",
  organism = NULL,
  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	species of the Entrez gene IDs. If omitted, it is inferred from 'ont'.
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

doseInterpretResult *Construct a canonical interpretation result*

Description

Construct a canonical interpretation result

Usage

```
doseInterpretResult(
  result = NULL,
  evidence = NULL,
  query = NULL,
  sources = NULL,
  parameters = list(),
  explanation = list()
)
```

Arguments

result	Ranked target rows.
evidence	Long-form evidence rows.
query	Query metadata rows.
sources	Source metadata rows.
parameters	Named list of analysis parameters.
explanation	Named list of explanation output.

Value

A validated ‘doseInterpretResult‘ object.

---

doseInterpretResult-class *Canonical interpretation result class*

Description

‘doseInterpretResult‘ is the single public result contract for disease interpretation workflows in DOSE.

Slots

`result` A data.frame of ranked interpretation targets.

`evidence` A data.frame of long-form evidence rows linked to targets.

`query` A data.frame describing the input query or queries.

`sources` A data.frame describing source versions and checksums.

`parameters` A named list of analysis parameters.

`explanation` A named list storing optional explanation output.

---

doseSim *doseSim*

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 *Shared parameters for DOSE functions*

Description

Shared parameters for DOSE functions

Arguments

gene	a vector of entrez gene id
organism	species of the input Entrez gene IDs. Use "hsa" for human or "mm" for mouse. Common aliases such as "human", "mouse", and "mmu" are accepted. If omitted, the species is inferred from ‘ont‘.
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"
seed	random seed for reproducibility, set to a number (or TRUE to use a fixed default seed) to make the result reproducible, or FALSE (default) to draw a random seed on each run, so results may vary between runs. The underlying permutation engine uses its own RNG seeded with this value; see `enrichit::gsea()` for details.

`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 = NULL,
  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	species of the input Entrez gene IDs. Use "hsa" for human or "mm" for mouse. Common aliases such as "human", "mouse", and "mmu" are accepted. If omitted, the species is inferred from 'ont'.
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 = enrichDO(gene, pvalueCutoff=0.05)
 summary(yy)
```

---

enrichNCG *enrichNCG*

Description

Enrichment analysis based on the Network of Cancer Genes database (https://www.network-cancer-genes.org/)

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

`evidence` *Filter evidence rows from a canonical interpretation result*

Description

Filter evidence rows from a canonical interpretation result

Usage

```
evidence(x, target_id = NULL, type = NULL, source = NULL, direction = NULL)

## S4 method for signature 'doseInterpretResult'
evidence(x, target_id = NULL, type = NULL, source = NULL, direction = NULL)
```

Arguments

`x`	A ‘doseInterpretResult‘ object.
`target_id`	Optional target identifier to keep.
`type`	Optional evidence type to keep.
`source`	Optional evidence source to keep.
`direction`	Optional evidence direction to keep.

Value

A filtered evidence data.frame.

`explainDisease` *Add evidence-grounded explanation text to a canonical interpretation result*

Description

Add evidence-grounded explanation text to a canonical interpretation result

Usage

```
explainDisease(
  x,
  method = c("template", "llm"),
  max_features = 3L,
  provider = NULL,
  model = NULL,
  language = c("en", "zh"),
  style = c("brief", "research", "cautious"),
  generation_params = list(),
  llm_adapter = NULL
)
```

Arguments

x	A ‘doseInterpretResult‘ object.
method	Explanation method. ‘"template"‘ is available offline.
max_features	Maximum number of evidence features to mention per query.
provider	Optional provider name passed to the LLM adapter.
model	Optional model name passed to the LLM adapter.
language	Explanation language for the optional LLM adapter.
style	Explanation style for the optional LLM adapter.
generation_params	Optional generation parameters passed to the LLM adapter.
llm_adapter	Optional adapter function for deterministic testing or custom provider integration.

Value

A ‘doseInterpretResult‘ with the ‘explanation‘ slot populated.

---

gene2DO *convert Gene ID to DO Terms*

Description

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

Usage

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

Arguments

gene	entrez gene ID
organism	species of the Entrez gene ID. If omitted, it is inferred from ‘ont‘.
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 = NULL,
  measure = "Wang",
  combine = "BMA"
)
```

Arguments

geneID1	entrez gene vector
geneID2	entrez gene vector
ont	one of "HDO" and "MPO"
organism	species of the Entrez gene IDs. If omitted, it is inferred from 'ont'.
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)
```

gseDO *DO Gene Set Enrichment Analysis*

Description

perform gsea analysis

Usage

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

Arguments

geneList	order ranked geneList
ont	one of "HDO", "HPO" or "MPO"
organism	species of the input Entrez gene IDs. Use "hsa" for human or "mm" for mouse. Common aliases such as "human", "mouse", and "mmu" are accepted. If omitted, the species is inferred from ‘ont‘.
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)

seed	random seed for reproducibility, set to a number (or TRUE to use a fixed default seed) to make the result reproducible, or FALSE (default) to draw a random seed on each run, so results may vary between runs. The underlying permutation engine uses its own RNG seeded with this value; see `enrichit::gsea()` for details.
...	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,
  seed = FALSE,
  ...
)
```

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)
seed	random seed for reproducibility, set to a number (or TRUE to use a fixed default seed) to make the result reproducible, or FALSE (default) to draw a random seed on each run, so results may vary between runs. The underlying permutation engine uses its own RNG seeded with this value; see `enrichit::gsea()` for details.
...	other parameter

Value

gseaResult object

Author(s)

Guangchuang Yu

---

interpretDisease *High-level disease interpretation entry point*

Description

This function defines the public interpretation contract and performs validation before the first disease-ranking implementation is enabled.

Usage

```
interpretDisease(
  x,
  input = c("auto", "gene", "ranked_gene", "gene_set_list", "phenotype"),
  organism = c("human", "mouse"),
  target = c("disease", "mouse_model", "both"),
  ontology = "auto",
  id_type = "auto",
  orthology = c("one_to_one", "all"),
  explain = c("none", "template", "llm"),
  top = 50,
  ...
)
```

Arguments

x	Query input.
input	One of ‘"auto"‘, ‘"gene"‘, ‘"ranked_gene"‘, ‘"gene_set_list"‘, or ‘"phenotype"‘.
organism	One of ‘"human"‘ or ‘"mouse"‘.
target	One of ‘"disease"‘, ‘"mouse_model"‘, or ‘"both"‘.
ontology	Ontology to use. ‘"auto"‘ currently resolves to ‘"HDO"‘ for the supported human disease path.
id_type	Identifier type. The current implementation accepts ‘"auto"‘ and ‘"ENTREZID"‘.
orthology	One of ‘"one_to_one"‘ or ‘"all"‘.
explain	One of ‘"none"‘, ‘"template"‘, or ‘"llm"‘.
top	Maximum number of targets to return.
...	Additional arguments passed to the underlying enrichment functions. Cross-species ‘target = "mouse_model"‘ and ‘target = "both"‘ also require ‘disease_model_file‘, ‘homology_file‘, and ‘gene_pheno_file‘.

Value

A ‘doseInterpretResult‘ once the ranking path is implemented.

---

mclusterSim *mclusterSim*

Description

Pairwise semantic similarity for a list of gene clusters

Usage

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

Arguments

clusters	A list of gene clusters
ont	one of "HDO", "HPO" and "MPO"
organism	species of the Entrez gene IDs. If omitted, it is inferred from ‘ont‘.
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)
```

---

rankMouseModels *Rank mouse disease models from human query genes*

Description

Rank mouse disease models from human query genes

Usage

```
rankMouseModels(
  gene,
  disease_model_file,
  homology_file,
  gene_pheno_file,
  top = 50
)
```

Arguments

gene	A character vector of human Entrez gene IDs.
disease_model_file	Path to an 'MGI_DiseaseMouseModel.rpt'-format file.
homology_file	Path to a 'HOM_MouseHumanSequence.rpt'-format file.
gene_pheno_file	Path to an 'MGI_GenePheno.rpt'-format file.
top	Maximum number of mouse-model targets to return.

Value

A 'doseInterpretResult' with 'target_type = "mouse_model"'.

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()
GOSemSim get_organism(), set_auto_update()

show,doseInterpretResult-method *Show a canonical interpretation result*

Description

Show a canonical interpretation result

Usage

```
## S4 method for signature 'doseInterpretResult'
show(object)
```

Arguments

object	A ‘doseInterpretResult‘ object.

Value

The object is shown to the console and returned invisibly.

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,doseInterpretResult-method

Summarize a canonical interpretation result

Description

Summarize a canonical interpretation result

Usage

```
## S4 method for signature 'doseInterpretResult'
summary(object, ...)
```

Arguments

object	A ‘doseInterpretResult‘ object.
...	Reserved for generic compatibility.

Value

A named list summarizing queries, targets, evidence, and sources.

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

∗ **datasets**
DataSet, 5
∗ **internal**
DOSE-package, 2
reexports, 18
∗ **manip**
enrichDO, 8
gseDO, 13
gseNCG, 14

as.data.frame,doseInterpretResult-method, 3

clusterSim, 4
computeIC, 5

DataSet, 5
DOSE *(*DOSE-package*)*, 2
DOSE-package, 2
dose_params, 7
doseInterpretResult, 5
doseInterpretResult-class, 6
doseSim, 6
doSim *(*doseSim*)*, 6

enrichDO, 8
enrichNCG, 9
evidence, 10
evidence,doseInterpretResult-method *(*evidence*)*, 10
explainDisease, 10

facet_grid *(*reexports*)*, 18
facet_grid(), *18*

gene2DO, 11
geneList *(*DataSet*)*, 5
geneSim, 12
get_organism *(*reexports*)*, 18
get_organism(), *18*
gseDO, 13
gseNCG, 14

interpretDisease, 15

mclusterSim, 16

rankMouseModels, 17
reexports, 18

set_auto_update *(*reexports*)*, 18
set_auto_update(), *18*
show,doseInterpretResult-method, 18
simplot, 18
summary,doseInterpretResult-method, 19

theme_dose, 20