

# Package ‘EnrichDO’

December 18, 2024

**Type** Package

**Title** a Global Weighted Model for Disease Ontology Enrichment Analysis

**Version** 1.1.1

## Description

To implement disease ontology (DO) enrichment analysis, this package is designed and presents a double weighted model based on the latest annotations of the human genome with DO terms, by integrating the DO graph topology on a global scale. This package exhibits high accuracy that it can identify more specific DO terms, which alleviates the over enriched problem. The package includes various statistical models and visualization schemes for discovering the associations between genes and diseases from biological big data.

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**Imports** BiocGenerics, Rgraphviz, clusterProfiler, hash, S4Vectors,  
dplyr, ggplot2, graph, magrittr, methods, pheatmap, graphics,  
utils, purrr, tidyr, stats

**biocViews** Annotation, Visualization, GeneSetEnrichment, Software

**Depends** R (>= 4.0.0)

**RoxygenNote** 7.3.1

**Suggests** knitr, rmarkdown, org.Hs.eg.db, testthat (>= 3.0.0),  
BiocStyle

**VignetteBuilder** knitr

**Config/testthat/edition** 3

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

**git\_branch** devel

**git\_last\_commit** ef3d0ef

**git\_last\_commit\_date** 2024-11-27

**Repository** Bioconductor 3.21

**Date/Publication** 2024-12-18

**Author** Liang Cheng [aut],  
Haixiu Yang [aut],  
Hongyu Fu [cre]

**Maintainer** Hongyu Fu <2287531995@qq.com>

## Contents

EnrichDO-package . . . . .	2
convDraw . . . . .	3
doEnrich . . . . .	3
dotermgenes . . . . .	5
doterms . . . . .	5
drawBarGraph . . . . .	6
drawGraphViz . . . . .	7
drawHeatmap . . . . .	8
drawPointGraph . . . . .	9
EnrichResult-class . . . . .	9
show,EnrichResult-method . . . . .	10
showDoTerms . . . . .	11
TermStruct . . . . .	11
writeResult . . . . .	12

**Index** **13**

---

EnrichDO-package	<i>EnrichDO Enrichment analyses including a variety of statistical models and visualization schemes for discovering the disease-gene relationship under biological big data.</i>
------------------	--

---

## Description

To implement disease ontology (DO) enrichment analysis, this package is designed and presents a double weighted model based on the latest annotations of the human genome with DO terms, by integrating the DO graph topology on a global scale. This package exhibits high accuracy that it can identify more specific DO terms, which alleviates the over enriched problem. The package includes various statistical models and visualization schemes for discovering the associations between genes and diseases from biological big data.

## Author(s)

Liang cheng, Haixiu Yang, Hongyu Fu

Maintainer: Haixiu Yang <yanghaixiu@ems.hrbmu.edu.cn>

---

`convDraw`*convDraw*

---

**Description**

using the result of writeResult for convenience drawing.

**Usage**

```
convDraw(resultDO)
```

**Arguments**

resultDO            a data frame of enrichment result

**Value**

DataFrame

**Author(s)**

Haixiu Yang

**Examples**

```
##Draw from writeResult output files
#Firstly, read the writeResult output file,using the following two lines
data <- read.delim(file.path(system.file('examples', package = 'EnrichDO'), 'result.txt'))
enrich <- convDraw(resultDO = data)
#then, Use the drawing function you need
drawGraphViz(enrich=enrich)    #Tree diagram
drawPointGraph(enrich=enrich) #Bubble diagram
drawBarGraph(enrich=enrich)   #Bar plot
```

---

`doEnrich`*doEnrich*

---

**Description**

given an array of human protein-genes with NCBI ENTREZID format, this function combines topological properties of the disease ontology structure for enrichment analysis.

**Usage**

```
doEnrich(
  interestGenes,
  test = c("hypergeomTest", "fisherTest", "binomTest", "chisqTest", "logoddTest"),
  method = c("BH", "holm", "hochberg", "hommel", "bonferroni", "BY", "fdr", "none"),
  m = 1,
  maxGsize = 5000,
  minGsize = 5,
  traditional = FALSE,
  delta = 0.01,
  penalize = TRUE,
  allDOTerms = FALSE
)
```

**Arguments**

interestGenes	a vector of gene IDs. The interest gene sets should be protein-coding genes, using the ENTREZID format from NCBI.
test	One of 'fisherTest', 'hypergeomTest', 'binomTest', 'chisqTest' and 'logoddTest' statistical model. Default is hypergeomTest.
method	One of 'holm', 'hochberg', 'hommel', 'bonferroni', 'BH', 'BY', 'fdr' and 'none', for P value correction.
m	Set the maximum number of ancestor layers for ontology enrichment. Default is layer 1.
maxGsize	indicates that doterms with more annotation genes than maxGsize are ignored, and the P value of these doterms is set to 1.
minGsize	indicates that doterms with less annotation genes than minGsize are ignored, and the P value of these doterms is set to 1.
traditional	a logical variable, TRUE for traditional enrichment analysis, FALSE for enrichment analysis with weights. Default is FALSE.
delta	Set the threshold of nodes, if the p value of doterm is greater than delta, the nodes are not significant, and these nodes are not weighted. Default is 0.01.
penalize	Logical value, used to alleviate the impact of different magnitudes of p-values, default value is TRUE. When set to FALSE, the degree of reduction in weight for non-significant nodes is decreased.
allDOTerms	Logical value, whether to store all doterms in EnrichResult, defaults is FALSE (only significant nodes are retained).

**Value**

A EnrichResult instance.

**Author(s)**

Haixiu Yang

**Examples**

```
##Input data case
#the inputdata_demo variable stores validated protein-coding genes associated with Alzheimer's disease.
Alzheimer <- read.delim(file.path(system.file('extdata', package='EnrichDO'), 'Alzheimer_curated.csv'), header =
inputdata_demo <- Alzheimer[,1]
##doEnrich case
#The enrichment results were obtained by using demo.data
demo.data <- c(1636,351,102,2932,3077,348,4137,54209)
demo_result <- doEnrich(interestGenes=demo.data,maxGsize = 100, minGsize=10)
```

---

dotermgenes	<i>All DO term annotated genes.</i>
-------------	-------------------------------------

---

**Description**

A dataset includes 15106 genes.

**Usage**

dotermgenes

**Format**

An character array with 15106 elements:

---

doterms	<i>Detailed annotation information for 4831 DO terms.</i>
---------	---

---

**Description**

A dataset includes 4831 DO terms of hierarchical information, annotated gene information, and weight information

**Usage**

doterms

**Format**

A data frame with 4813 rows and 10 variables:

**DOID** the DOterm ID on enrichment

**level** the hierarchy of the DOterm in the DAG graph

**gene.arr** all genes related to the DOterm

**weight.arr** gene weights in each node

**parent.arr** the parent node of the DOterm  
**parent.len** the number of parent.arr  
**child.arr** child nodes of the DOterm  
**child.len** the number of child.arr  
**gene.len** the number of all genes related to the DOterm  
**DOTerm** the standard name of the DOterm

---

drawBarGraph

*drawBarGraph*

---

### Description

The enrichment results are shown in a bar chart

### Usage

```
drawBarGraph(EnrichResult = NULL, enrich = NULL, n = 10, delta = 1e-15)
```

### Arguments

EnrichResult	the EnrichResult object
enrich	a data frame of enrichment result
n	number of bars
delta	the threshold of P value

### Value

bar graph

### Author(s)

Haixiu Yang

### Examples

```
demo.data <- c(1636,351,102,2932,3077,348,4137,54209)
sample1 <- doEnrich(interestGenes=demo.data,maxGsize = 100, minGsize=10)
drawBarGraph(EnrichResult=sample1, n=10, delta=0.05)
```

---

drawGraphViz	<i>drawGraphViz</i>
--------------	---------------------

---

**Description**

the enrichment results are shown in a tree diagram

**Usage**

```
drawGraphViz(  
  EnrichResult = NULL,  
  enrich = NULL,  
  n = 10,  
  labelfontsize = 14,  
  numview = TRUE,  
  pview = TRUE  
)
```

**Arguments**

EnrichResult	the EnrichResult object
enrich	a data frame of the enrichment result
n	the number of most significant nodes
labelfontsize	the font size of nodes
numview	Displays the number of intersections between the interest set and each doterm.
pview	Displays the P value for each dotrem.

**Value**

tree diagram

**Author(s)**

Haixiu Yang

**Examples**

```
demo.data <- c(1636,351,102,2932,3077,348,4137,54209)  
sample5 <- doEnrich(interestGenes=demo.data,maxGsize = 100, minGsize=10)  
drawGraphViz(EnrichResult =sample5)  
  
#The p-value and the number of intersections are not visible  
drawGraphViz(EnrichResult=sample5, numview = FALSE, pview = FALSE)
```

---

drawHeatmap	<i>drawHeatmap</i>
-------------	--------------------

---

### Description

The top DOID\_n nodes in the enrichment results showed the top gene\_n genes with the highest weight sum.

### Usage

```
drawHeatmap(  
  interestGenes,  
  EnrichResult = NULL,  
  DOID_n = 10,  
  gene_n = 50,  
  fontsize_row = 10,  
  readable = TRUE,  
  ...  
)
```

### Arguments

interestGenes	A collection of interest genes in vector form
EnrichResult	the EnrichResult object
DOID_n	There are DOID_n nodes with the highest significance in the enrichment results.
gene_n	Among the selected DOID_n nodes, the top gene_n genes with the highest weight sum are selected to show.
fontsize_row	Set the font size of the gene tag.
readable	Logical value that controls whether the gene tag is in symbol format
...	Other parameters in the pheatmap function also apply.

### Value

heat map

### Author(s)

Haixiu Yang

### Examples

```
demo.data <- c(1636,351,102,2932,3077,348,4137,54209)  
sample6 <- doEnrich(interestGenes=demo.data,maxGsize = 100, minGsize=10)  
drawHeatmap(interestGenes=demo.data, EnrichResult = sample6, gene_n = 10)
```



---

drawPointGraph	<i>drawPointGraph</i>
----------------	-----------------------

---

**Description**

The enrichment results are shown in a scatter plot

**Usage**

```
drawPointGraph(EnrichResult = NULL, enrich = NULL, n = 10, delta = 1e-15)
```

**Arguments**

EnrichResult	the EnrichResult object
enrich	a data frame of enrichment result.
n	number of points.
delta	the threshold of P value.

**Value**

scatter graph

**Author(s)**

Haixiu Yang

**Examples**

```
demo.data <- c(1636,351,102,2932,3077,348,4137,54209)
sample2 <- doEnrich(interestGenes=demo.data,maxGsize = 100, minGsize=10)
drawPointGraph(EnrichResult=sample2, n=10, delta=0.05)
```

---

EnrichResult-class	<i>Class 'EnrichResult' This class represents the result of enrich analysis</i>
--------------------	---

---

**Description**

Class 'EnrichResult' This class represents the result of enrich analysis

**Slots**

enrich a data frame of enrichment result  
test Statistical test  
method Multiple test correction methods  
m the maximum number of ancestor layers for ontology enrichment  
maxGsize The maximum number of DOTerm genes in enrichment analysis  
minGsize The minimum number of DOTerm genes in enrichment analysis  
traditional Indicates whether the traditional ORA method is used  
delta The highest p-value of significance for each node  
penalize Whether to use penalty function in enrichment analysis  
interestGenes A valid interest gene set

**Author(s)**

Haixiu Yang

---

show,EnrichResult-method

*show method*

---

**Description**

show method for EnrichResult instance

**Usage**

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

**Arguments**

object A EnrichResult instance.

**Value**

print info

**Author(s)**

Haixiu Yang

---

showDoTerms	<i>showDoTerms</i>
-------------	--------------------

---

**Description**

show DOterms

**Usage**

```
showDoTerms(doterms = doterms)
```

**Arguments**

doterms            a data frame of DOterms.

**Value**

text

**Author(s)**

Haixiu Yang

**Examples**

```
showDoTerms(doterms)
```

---

TermStruct	<i>Enrich_internal</i>
------------	------------------------

---

**Description**

Internal calculation of enrichment analysis

**Usage**

```
TermStruct(resultDO)
```

**Arguments**

resultDO            Receives the file output by the writeResult function, which is used to visually display the enrichment results (without running the enrichment operation again).

**Value**

A EnrichResult instance.

**Author(s)**

Haixiu Yang

---

`writeResult`*writeResult*

---

**Description**

Output enrichment result as text

**Usage**

```
writeResult(EnrichResult = NULL, file, Q = 1, P = 1)
```

**Arguments**

EnrichResult	the EnrichResult object
file	the address and name of the output file.
Q	Output only doterm information with p.adjust values less than or equal to Q.
P	Output only doterm information with p values less than or equal to P.

**Value**

text

**Author(s)**

Haixiu Yang

**Examples**

```
demo.data <- c(1636,351,102,2932,3077,348,4137,54209)
sample4 <- doEnrich(interestGenes=demo.data,maxGsize = 100, minGsize=10)
writeResult(EnrichResult=sample4, file=file.path(tempdir(), 'result.txt'))
```

# Index

- \* **classes**
  - EnrichResult-class, [9](#)
- \* **datasets**
  - dotermgenes, [5](#)
  - doterms, [5](#)
- \* **package**
  - EnrichDO-package, [2](#)

[convDraw](#), [3](#)

[doEnrich](#), [3](#)

[dotermgenes](#), [5](#)

[doterms](#), [5](#)

[drawBarGraph](#), [6](#)

[drawGraphViz](#), [7](#)

[drawHeatmap](#), [8](#)

[drawPointGraph](#), [9](#)

[EnrichDO \(EnrichDO-package\)](#), [2](#)

[EnrichDO-package](#), [2](#)

[EnrichResult-class](#), [9](#)

[show](#), [EnrichResult-method](#), [10](#)

[showDoTerms](#), [11](#)

[TermStruct](#), [11](#)

[writeResult](#), [12](#)