Package 'clusterProfiler'

October 12, 2016

Type Package

Title statistical analysis and visualization of functional profiles
for genes and gene clusters

Version 3.0.5

Author Guangchuang Yu <guangchuangyu@gmail.com> with contributions from
Li-Gen Wang and Giovanni Dall'Olio.

Maintainer Guangchuang Yu <guangchuangyu@gmail.com>

Description This package implements methods to analyze and visualize
functional profiles (GO and KEGG) of gene and gene clusters.

Depends R (>= 3.2.0), DOSE (>= 2.9.0)

Imports AnnotationDbi, ggplot2, GO.db, GOSemSim, GSEABase, IRanges,
magrittr, methods, plyr, qvalue, stats, stats4, S4Vectors,
tidyr, topGO, utils

Suggests AnnotationHub, BiocStyle, KEGG.db, knitr, org.Hs.eg.db,
pathview, ReactomePA

VignetteBuilder knitr

License Artistic-2.0

URL http://guangchuangyu.github.io/clusterProfiler

BugReports https://github.com/GuangchuangYu/clusterProfiler/issues

biocViews Annotation, Clustering, GeneSetEnrichment, GO, KEGG,
MultipleComparison, Pathways, Reactome, Visualization

RoxygenNote 5.0.1

NeedsCompilation no

R topics documented:

clusterProfiler-package

> *statistical analysis and visualization of functional profiles for genes and gene clusters The package implements methods to analyze and visualize functional profiles of gene and gene clusters.*

Description

This package is designed to compare gene clusters functional profiles.

Details

Package:	clusterProfiler
Type:	Package
Version:	1.9.
Date:	06-13-2013
biocViews:	GO, Clustering, Visulization
Depends:	AnnotationDbi, GO.db, org.Hs.eg.db, ggplot2, plyr, methods
Suggests:	GOSemSim
License:	Artistic-2.0

Author(s)

Guangchuang Yu <guangchuangyu@gmail.com>

Maintainer: Guangchuang Yu <guangchuangyu@gmail.com>

See Also

compareClusterResult, groupGOResult enrichResult

---

bitr	*bitr*

Description

Biological Id TRanslator

Usage

```
bitr(geneID, fromType, toType, OrgDb, drop = TRUE)
```

Arguments

geneID	input gene id
fromType	input id type
toType	output id type
OrgDb	annotation db
drop	drop NA or not

Value

data.frame

Author(s)

Guangchuang Yu

---

bitr_kegg   *bitr_kegg*

Description

convert biological ID using KEGG API

Usage

```
bitr_kegg(geneID, fromType, toType, organism, drop = TRUE)
```

Arguments

geneID	input gene id
fromType	input id type
toType	output id type
organism	supported organism, can be search using search_kegg_organism function
drop	drop NA or not

Value

data.frame

Author(s)

Guangchuang Yu

---

buildGOmap   *buildGOmap*

Description

building GO mapping files

Usage

```
buildGOmap(gomap)
```

Arguments

gomap	data.frame with two columns of GO and gene ID

Details

provided by a data.frame of GO (column 1) and gene (column 2) direct annotation this function will building gene to GO and GO to gene mapping, with directly and undirectly (ancestor GO term) annotation.

Value

data.frame, GO annotation with indirect annotation

Author(s)

Yu Guangchuang

---

compareCluster   *Compare gene clusters functional profile*

Description

Given a list of gene set, this function will compute profiles of each gene cluster.

Usage

```
compareCluster(geneClusters, fun = "enrichGO", data = "", ...)
```

Arguments

geneClusters	a list of entrez gene id. Alternatively, a formula of type Entrez~group
fun	One of "groupGO", "enrichGO", "enrichKEGG", "enrichDO" or "enrichPathway" .
data	if geneClusters is a formula, the data from which the clusters must be extracted.
...	Other arguments.

Value

A clusterProfResult instance.

Author(s)

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

See Also

compareClusterResult-class, groupGO enrichGO

Examples

```
data(gcSample)
xx <- compareCluster(gcSample, fun="enrichKEGG",
                     organism="hsa", pvalueCutoff=0.05)
summary(xx)
# plot(xx, type="dot", caption="KEGG Enrichment Comparison")

## formula interface
mydf <- data.frame(Entrez=c('1', '100', '1000', '100101467',
                            '100127206', '100128071'),
                   group = c('A', 'A', 'A', 'B', 'B', 'B'),
                   othergroup = c('good', 'good', 'bad', 'bad', 'good', 'bad'))
xx.formula <- compareCluster(Entrez~group, data=mydf, fun='groupGO', OrgDb='org.Hs.eg.db')
summary(xx.formula)

## formula interface with more than one grouping variable
xx.formula.twogroups <- compareCluster(Entrez~group+othergroup, data=mydf, fun='groupGO', OrgDb='org.Hs.eg.db')
summary(xx.formula.twogroups)
```

---

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

.call function call

Author(s)

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

See Also

[groupGOResult](#) [enrichResult](#) [compareCluster](#)

DataSet	*Datasets gcSample contains a sample of gene clusters.*

Description

Datasets gcSample contains a sample of gene clusters.

Datasets kegg_species contains kegg species information

dotplot,compareClusterResult-method
dotplot

Description

dot plot method

Usage

```
## S4 method for signature 'compareClusterResult'
dotplot(object, x = ~Cluster,
  colorBy = "p.adjust", showCategory = 5, by = "geneRatio",
  includeAll = TRUE, font.size = 12, title = "")
```

Arguments

object	compareClusterResult object
x	x variable
colorBy	one of pvalue or p.adjust
showCategory	category numbers
by	one of geneRatio, Percentage or count
includeAll	logical
font.size	font size
title	figure title

---

download.KEGG   *download.KEGG*

Description

download the latest version of KEGG pathway

Usage

```
download.KEGG(species, KEGG_Type = "KEGG", keyType = "kegg")
```

Arguments

species   species

KEGG_Type   one of 'KEGG' or 'MKEGG'

keyType   supported keyType, see bitr_kegg

Value

list

Author(s)

Guangchuang Yu

---

dropGO   *dropGO*

Description

drop GO term of specific level or specific terms (mostly too general).

Usage

```
dropGO(x, level = NULL, term = NULL)
```

Arguments

x   an instance of 'enrichResult' or 'compareClusterResult'

level   GO level

term   GO term

Value

modified version of x

Author(s)

Guangchuang Yu

---

enrichDAVID   *enrichDAVID*

Description

enrichment analysis by DAVID

Usage

```
enrichDAVID(gene, idType = "ENTREZ_GENE_ID", listType = "Gene",
  minGSSize = 10, maxGSSize = 500, annotation = "GOTERM_BP_FAT",
  pvalueCutoff = 0.05, pAdjustMethod = "BH", qvalueCutoff = 0.2,
  species = NA, david.user)
```

Arguments

gene	input gene
idType	id type
listType	list Type
minGSSize	minimal size of genes annotated for testing
maxGSSize	maximal size of genes annotated for testing
annotation	david annotation
pvalueCutoff	pvalueCutoff
pAdjustMethod	one of "BH" and "bonferroni"
qvalueCutoff	qvalutCutoff
species	species
david.user	david user

Value

A enrichResult instance

Author(s)

Guangchuang Yu

enricher	*enricher*

Description

A universal enrichment analyzer

Usage

```
enricher(gene, pvalueCutoff = 0.05, pAdjustMethod = "BH", universe,
  minGSSize = 10, maxGSSize = 500, qvalueCutoff = 0.2, TERM2GENE,
  TERM2NAME = NA)
```

Arguments

gene	a vector of 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 for testing
maxGSSize	maximal size of genes annotated for testing
qvalueCutoff	qvalue cutoff
TERM2GENE	user input annotation of TERM TO GENE mapping, a data.frame of 2 column with term and gene
TERM2NAME	user input of TERM TO NAME mapping, a data.frame of 2 column with term and name

Value

A enrichResult instance

Author(s)

Guangchuang Yu

---

enrichGO   *GO Enrichment Analysis of a gene set. Given a vector of genes, this*
                                 function will return the enrichment GO categories after FDR control.

Description

GO Enrichment Analysis of a gene set. Given a vector of genes, this function will return the enrichment GO categories after FDR control.

Usage

```
enrichGO(gene, OrgDb, keytype = "ENTREZID", ont = "MF",
  pvalueCutoff = 0.05, pAdjustMethod = "BH", universe, qvalueCutoff = 0.2,
  minGSSize = 10, maxGSSize = 500, readable = FALSE)
```

Arguments

gene   a vector of entrez gene id.

OrgDb   OrgDb

keytype   keytype of input gene

ont   One of "MF", "BP", and "CC" subontologies.

pvalueCutoff   Cutoff value of pvalue.

pAdjustMethod   one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"

universe   background genes

qvalueCutoff   qvalue cutoff

minGSSize   minimal size of genes annotated by Ontology term for testing.

maxGSSize   maximal size of genes annotated for testing

readable   whether mapping gene ID to gene Name

Value

A enrichResult instance.

Author(s)

Guangchuang Yu <http://ygc.name>

See Also

[enrichResult-class](#), [compareCluster](#)

Examples

```
## Not run:
data(gcSample)
yy <- enrichGO(gcSample[[1]], 'org.Hs.eg.db', ont="BP", pvalueCutoff=0.01)
head(summary(yy))
plot(yy)

## End(Not run)
```

---

enrichKEGG	*KEGG Enrichment Analysis of a gene set. Given a vector of genes, this function will return the enrichment KEGG categories with FDR control.*

Description

KEGG Enrichment Analysis of a gene set. Given a vector of genes, this function will return the enrichment KEGG categories with FDR control.

Usage

```
enrichKEGG(gene, organism = "hsa", keyType = "kegg", pvalueCutoff = 0.05,
  pAdjustMethod = "BH", universe, minGSSize = 10, maxGSSize = 500,
  qvalueCutoff = 0.2, use_internal_data = FALSE)
```

Arguments

gene	a vector of entrez gene id.
organism	supported organism listed in 'http://www.genome.jp/kegg/catalog/org_list.html'
keyType	one of "kegg", 'ncbi-geneid', 'ncib-proteinid' and 'uniprot'
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 genes annotated for testing
qvalueCutoff	qvalue cutoff
use_internal_data	
	logical, use KEGG.db or latest online KEGG data

Value

A enrichResult instance.

Author(s)

Guangchuang Yu <http://ygc.name>

See Also

[enrichResult-class](), [compareCluster]()

Examples

```
data(gcSample)
yy = enrichKEGG(gcSample[[5]], pvalueCutoff=0.01)
head(summary(yy))
#plot(yy)
```

---

enrichMKEGG   *KEGG Module Enrichment Analysis of a gene set. Given a vector of*
                           genes, this function will return the enrichment KEGG Module cate-
                           gories with FDR control.

Description

KEGG Module Enrichment Analysis of a gene set. Given a vector of genes, this function will return the enrichment KEGG Module categories with FDR control.

Usage

```
enrichMKEGG(gene, organism = "hsa", keyType = "kegg", pvalueCutoff = 0.05,
  pAdjustMethod = "BH", universe, minGSSize = 10, maxGSSize = 500,
  qvalueCutoff = 0.2)
```

Arguments

gene	a vector of entrez gene id.
organism	supported organism listed in 'http://www.genome.jp/kegg/catalog/org_list.html'
keyType	one of "kegg", 'ncbi-geneid', 'ncib-proteinid' and 'uniprot'
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 genes annotated for testing
qvalueCutoff	qvalue cutoff

Value

A enrichResult instance.

---

fortify.compareClusterResult
fortify

Description

convert compareClusterResult to a data.frame that ready for plot

Usage

```
## S3 method for class 'compareClusterResult'
fortify(model, data, showCategory = 5,
  by = "geneRatio", includeAll = TRUE)
```

Arguments

model	compareClusterResult object
data	not use here
showCategory	category numbers
by	one of geneRatio, Percentage or count
includeAll	logical

Value

data.frame

Author(s)

Guangchuang Yu

---

getGOLevel   *get GOIDs at a specific level*

Description

query GOIDs at a specific level.

Usage

```
getGOLevel(ont, level)
```

Arguments

ont	Ontology
level	GO level

Value

a vector of GOIDs

Author(s)

Guangchuang Yu <http://ygc.name>

Gff2GeneTable   *Gff2GeneTable*

Description

read GFF file and build gene information table

Usage

```
Gff2GeneTable(gffFile, compress = TRUE)
```

Arguments

gffFile   GFF file

compress   compress file or not

Details

given a GFF file, this function extracts information from it and save it in working directory

Value

file save.

Author(s)

Yu Guangchuang

---

go2ont *go2ont*

Description

convert goid to ontology (BP, CC, MF)

Usage

```
go2ont(goid)
```

Arguments

goid   a vector of GO IDs

Value

data.frame

Author(s)

Guangchuang Yu

---

go2term *go2term*

Description

convert goid to descriptive term

Usage

```
go2term(goid)
```

Arguments

goid   a vector of GO IDs

Value

data.frame

Author(s)

Guangchuang Yu

---

gofilter	*gofilter*

Description

filter GO enriched result at specific level

Usage

```
gofilter(x, level = 4)
```

Arguments

x	output from enrichGO or compareCluster
level	GO level

Value

updated object

Author(s)

Guangchuang Yu

---

groupGO	*Functional Profile of a gene set at specific GO level. Given a vector of genes, this function will return the GO profile at a specific level.*

Description

Functional Profile of a gene set at specific GO level. Given a vector of genes, this function will return the GO profile at a specific level.

Usage

```
groupGO(gene, OrgDb, keytype = "ENTREZID", ont = "CC", level = 2,
  readable = FALSE)
```

Arguments

gene	a vector of entrez gene id.
OrgDb	OrgDb
keytype	keytype of input gene
ont	One of "MF", "BP", and "CC" subontologies.
level	Specific GO Level.
readable	if readable is TRUE, the gene IDs will mapping to gene symbols.

Value

A groupGOResult instance.

Author(s)

Guangchuang Yu http://ygc.name

See Also

groupGOResult-class, compareCluster

Examples

```
data(gcSample)
yy <- groupGO(gcSample[[1]], 'org.Hs.eg.db', ont="BP", level=2)
head(summary(yy))
#plot(yy)
```

---

groupGOResult-class   *Class "groupGOResult" This class represents the result of functional*
                         Profiles of a set of gene at specific GO level.

Description

Class "groupGOResult" This class represents the result of functional Profiles of a set of gene at specific GO level.

Slots

result GO classification result

ontology Ontology

level GO level

organism one of "human", "mouse" and "yeast"

gene Gene IDs

geneInCategory gene and category association

readable logical flag of gene ID in symbol or not.

Author(s)

Guangchuang Yu http://ygc.name

See Also

compareClusterResult compareCluster groupGO

GSEA   *GSEA*

Description

a universal gene set enrichment analysis tools

Usage

```
GSEA(geneList, exponent = 1, nPerm = 1000, minGSSize = 10,
  maxGSSize = 500, pvalueCutoff = 0.05, pAdjustMethod = "BH", TERM2GENE,
  TERM2NAME = NA, verbose = TRUE, seed = FALSE)
```

Arguments

geneList	order ranked geneList
exponent	weight of each step
nPerm	number of permutations
minGSSize	minimal size of each geneSet for analyzing
maxGSSize	maximal size of genes annotated for testing
pvalueCutoff	pvalue cutoff
pAdjustMethod	p value adjustment method
TERM2GENE	user input annotation of TERM TO GENE mapping, a data.frame of 2 column with term and gene
TERM2NAME	user input of TERM TO NAME mapping, a data.frame of 2 column with term and name
verbose	logical
seed	logical

Value

gseaResult object

Author(s)

Guangchuang Yu

---

gseGO   *gseGO*

Description

Gene Set Enrichment Analysis of Gene Ontology

Usage

```
gseGO(geneList, ont = "BP", OrgDb, keytype = "ENTREZID", exponent = 1,
  nPerm = 1000, minGSSize = 10, maxGSSize = 500, pvalueCutoff = 0.05,
  pAdjustMethod = "BH", verbose = TRUE, seed = FALSE)
```

Arguments

geneList	order ranked geneList
ont	one of "BP", "MF", "CC" or "GO"
OrgDb	OrgDb
keytype	keytype of gene
exponent	weight of each step
nPerm	permutation numbers
minGSSize	minimal size of each geneSet for analyzing
maxGSSize	maximal size of genes annotated for testing
pvalueCutoff	pvalue Cutoff
pAdjustMethod	pvalue adjustment method
verbose	print message or not
seed	logical

Value

gseaResult object

Author(s)

Yu Guangchuang

---

gseKEGG   *gseKEGG*

Description

Gene Set Enrichment Analysis of KEGG

Usage

```
gseKEGG(geneList, organism = "hsa", keyType = "kegg", exponent = 1,
  nPerm = 1000, minGSSize = 10, maxGSSize = 500, pvalueCutoff = 0.05,
  pAdjustMethod = "BH", verbose = TRUE, use_internal_data = FALSE,
  seed = FALSE)
```

Arguments

geneList	order ranked geneList
organism	supported organism listed in 'http://www.genome.jp/kegg/catalog/org_list.html'
keyType	one of "kegg", 'ncbi-geneid', 'ncib-proteinid' and 'uniprot'
exponent	weight of each step
nPerm	permutation numbers
minGSSize	minimal size of each geneSet for analyzing
maxGSSize	maximal size of genes annotated for testing
pvalueCutoff	pvalue Cutoff
pAdjustMethod	pvalue adjustment method
verbose	print message or not
use_internal_data	
	logical, use KEGG.db or latest online KEGG data
seed	logical

Value

gseaResult object

Author(s)

Yu Guangchuang

---

gseMKEGG   *gseMKEGG*

Description

Gene Set Enrichment Analysis of KEGG Module

Usage

```
gseMKEGG(geneList, organism = "hsa", keyType = "kegg", exponent = 1,
  nPerm = 1000, minGSSize = 10, maxGSSize = 500, pvalueCutoff = 0.05,
  pAdjustMethod = "BH", verbose = TRUE, seed = FALSE)
```

Arguments

geneList	order ranked geneList
organism	supported organism listed in 'http://www.genome.jp/kegg/catalog/org_list.html'
keyType	one of "kegg", 'ncbi-geneid', 'ncib-proteinid' and 'uniprot'
exponent	weight of each step
nPerm	permutation numbers
minGSSize	minimal size of each geneSet for analyzing
maxGSSize	maximal size of genes annotated for testing
pvalueCutoff	pvalue Cutoff
pAdjustMethod	pvalue adjustment method
verbose	print message or not
seed	logical

Value

gseaResult object

Author(s)

Yu Guangchuang

---

idType *idType*

Description

list ID types supported by annoDb

Usage

```
idType(OrgDb = "org.Hs.eg.db")
```

Arguments

OrgDb   annotation db

Value

character vector

Author(s)

Guangchuang Yu

---

merge_result *merge_result*

Description

merge a list of enrichResult objects to compareClusterResult

Usage

```
merge_result(enrichResultList)
```

Arguments

enrichResultList
                  a list of enrichResult objects

Value

a compareClusterResult instance

Author(s)

Guangchuang Yu

---

plot   *plot method*

Description

plot method generics

Usage

```
## S4 method for signature 'compareClusterResult,ANY'
plot(x, type = "dot",
  colorBy = "p.adjust", showCategory = 5, by = "geneRatio",
  includeAll = TRUE, font.size = 12, title = "")
```

Arguments

x	compareClusterResult object
type	one of bar or dot
colorBy	one of pvalue or p.adjust
showCategory	category numbers
by	one of geneRatio, Percentage or count
includeAll	logical
font.size	font size
title	figure title
...	Additional argument list

Value

plot

Author(s)

Guangchuang Yu <http://ygc.name>

plotGOgraph   *plotGOgraph*

Description

plot GO graph

Usage

```
plotGOgraph(x, firstSigNodes = 10, useInfo = "all", sigForAll = TRUE,
  useFullNames = TRUE, ...)
```

Arguments

x	output of enrichGO or gseGO
firstSigNodes	number of significant nodes (retangle nodes in the graph)
useInfo	additional info
sigForAll	if TRUE the score/p-value of all nodes in the DAG is shown, otherwise only score will be shown
useFullNames	logical
...	additional parameter of showSigOfNodes, please refer to topGO

Value

GO DAG graph

Author(s)

Guangchuang Yu

See Also

showSigOfNodes

---

plotting.clusterProfile

plotting-clusterProfile

Description

Internal plot function for plotting compareClusterResult

Usage

```
plotting.clusterProfile(clProf.reshape.df, x = ~Cluster, type = "dot",
  colorBy = "p.adjust", by = "geneRatio", title = "", font.size = 12)
```

Arguments

clProf.reshape.df

               data frame of compareCluster result

x	x variable
type	one of dot and bar
colorBy	one of pvalue or p.adjust
by	one of percentage and count
title	graph title
font.size	graph font size

Value

ggplot object

Author(s)

Guangchuang Yu http://ygc.name

---

read.gmt   *read.gmt*

Description

parse gmt file to a data.frame

Usage

```
read.gmt(gmtfile)
```

Arguments

gmtfile   gmt file

Value

data.frame

Author(s)

Guangchuang Yu

---

search_kegg_organism   *search_kegg_organism*

Description

search kegg organism, listed in http://www.genome.jp/kegg/catalog/org_list.html

Usage

```
search_kegg_organism(str, by = "scientific_name", ignore.case = FALSE)
```

Arguments

str   string
by   one of 'kegg.code', 'scientific_name' and 'common_name'
ignore.case   TRUE or FALSE

Value

data.frame

Author(s)

Guangchuang Yu

---

simplify   *simplify method*

Description

simplify output from enrichGO by removing redundancy of enriched GO terms

simplify output from compareCluster by removing redundancy of enriched GO terms

Usage

```
## S4 method for signature 'enrichResult'
simplify(x, cutoff = 0.7, by = "p.adjust",
  select_fun = min, measure = "Rel")

## S4 method for signature 'compareClusterResult'
simplify(x, cutoff = 0.7, by = "p.adjust",
  select_fun = min, measure = "Rel")
```

Arguments

x	output of enrichGO
cutoff	similarity cutoff
by	feature to select representative term, selected by 'select_fun' function
select_fun	function to select feature passed by 'by' parameter
measure	method to measure similarity

Value

updated enrichResult object

updated compareClusterResult object

Author(s)

Guangchuang Yu

Guangchuang Yu

References

issue #28 https://github.com/GuangchuangYu/clusterProfiler/issues/28

---

viewKEGG | *viewKEGG function is for visualize KEGG pathways works with en-richResult object to visualize enriched KEGG pathway*

Description

viewKEGG function is for visualize KEGG pathways works with enrichResult object to visualize enriched KEGG pathway

Usage

```
viewKEGG(obj, pathwayID, foldChange, color.low = "green",
  color.high = "red", kegg.native = TRUE, out.suffix = "clusterProfiler")
```

Arguments

obj	enrichResult object
pathwayID	pathway ID or index
foldChange	fold change values
color.low	color of low foldChange genes
color.high	color of high foldChange genes
kegg.native	logical
out.suffix	suffix of output file

References

Luo et al. (2013) Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* (Oxford, England), 29:14 1830–1831, 2013. ISSN 1367-4803 http://bioinformatics.oxfordjournals.org/content/abstract/29/14/1830.abstract PMID: 23740750

Index