

Package ‘clusterProfiler’

October 8, 2015

Type Package

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

Version 2.2.7

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 (>= 3.1.0)

Imports DOSE, GOSemSim, AnnotationDbi, methods, stats4, plyr, ggplot2, GO.db, KEGGREST, magrittr, qvalue

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

VignetteBuilder knitr

License Artistic-2.0

URL <https://github.com/GuangchuangYu/clusterProfiler>

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

biocViews Clustering, GO, Pathways, Visualization, MultipleComparison, GeneSetEnrichment

NeedsCompilation no

R topics documented:

clusterProfiler-package	2
bitr	3
buildGOMap	4
buildKEGGmap	4
cnetplot	5
compareCluster	6

compareClusterResult-class	7
DataSet	7
download.KEGG	8
enrichDAVID	8
enricher	9
enrichGO	10
enrichKEGG	11
enrichMap	12
getGOLevel	13
Gff2GeneTable	13
groupGO	14
groupGOResult-class	15
GSEA	15
gseaplot	16
gseGO	17
gseKEGG	18
idType	19
plot	19
plotting.clusterProfile	20
viewKEGG	21

Index 22

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, Visualization
 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, annoDb, drop = TRUE)
```

Arguments

geneID	input gene id
fromType	input id type
toType	output id type
annoDb	annotation db
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 indirectly (ancestor GO term) annotation.

Value

GO annotation

Author(s)

Yu Guangchuang

 buildKEGGmap

buildKEGGmap

Description

build KEGG annotation files

Usage

```
buildKEGGmap(keggmap, id2name = NULL, organism)
```

Arguments

keggmap pathway to external ID
 id2name pathway id to pathway name
 organism organism

Author(s)

Guangchuang Yu

cnetplot

cnetplot

Description

category-gene-net plot

Usage

```
cnetplot(x, showCategory = 5, categorySize = "geneNum", foldChange = NULL,  
         fixed = TRUE, ...)
```

Arguments

<code>x</code>	enrichResult object
<code>showCategory</code>	number of category plotted
<code>categorySize</code>	one of geneNum or pvalue
<code>foldChange</code>	fold change of expression value
<code>fixed</code>	logical
<code>...</code>	additional parameter

Details

category gene association

Value

plot

Author(s)

ygc

compareCluster	<i>Compare gene clusters functional profile</i>
----------------	---

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 <code>Entrez~group</code>
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://ygc.name>

See Also

[compareClusterResult-class](#), [groupGO](#) [enrichGO](#)

Examples

```
data(gcSample)
xx <- compareCluster(gcSample, fun="enrichKEGG",
                    organism="human", 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')
summary(xx.formula)

## formula interface with more than one grouping variable
```

```
xx.formula.twogroups <- compareCluster(Entrez~group+othergroup, data=mydf, fun='groupGO')  
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

Author(s)

Guangchuang Yu <http://ygc.name>

See Also

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

DataSet

Datasets gcSample contains a sample of gene clusters.

Description

Datasets gcSample contains a sample of gene clusters.

download.KEGG	<i>download.KEGG</i>
---------------	----------------------

Description

download the latest version of KEGG pathway

Usage

```
download.KEGG(species)
```

Arguments

species	species
---------	---------

Value

list

Author(s)

Guangchuang Yu

enrichDAVID	<i>enrichDAVID</i>
-------------	--------------------

Description

enrichment analysis by DAVID

Usage

```
enrichDAVID(gene, idType = "ENTREZ_GENE_ID", listType = "Gene",
  minGSSize = 5, annotation = "GOTERM_BP_ALL", pvalueCutoff = 0.05,
  pAdjustMethod = "BH", qvalueCutoff = 0.2, species = NA,
  david.user = "clusterProfiler@hku.hk")
```

Arguments

gene	input gene
idType	id type
listType	list Type
minGSSize	minGSSize
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	<i>enricher</i>
----------	-----------------

Description

A universal enrichment analyzer

Usage

```
enricher(gene, pvalueCutoff = 0.05, pAdjustMethod = "BH", universe,
  minGSSize = 5, 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
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	<i>GO Enrichment Analysis of a gene set. Given a vector of genes, this function will return the enrichment GO categories after FDR control.</i>
----------	---

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, organism = "human", ont = "MF", pvalueCutoff = 0.05,
         pAdjustMethod = "BH", universe, qvalueCutoff = 0.2, minGSSize = 5,
         readable = FALSE)
```

Arguments

gene	a vector of entrez gene id.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coelicolor", "ecolik12", "ecsakai", "fly", "gondii", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast" and "zebrafish".
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.
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]], organism="human", ont="BP", pvalueCutoff=0.01)
head(summary(yy))
plot(yy)

## End(Not run)
```

enrichKEGG	<i>KEGG Enrichment Analysis of a gene set. Given a vector of genes, this function will return the enrichment KEGG categories with FDR control.</i>
------------	--

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 = "human", pvalueCutoff = 0.05,
  pAdjustMethod = "BH", universe, minGSSize = 5, qvalueCutoff = 0.2,
  readable = FALSE, use_internal_data = FALSE)
```

Arguments

gene	a vector of entrez gene id.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "ecoli12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast" and "zebrafish".
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.
qvalueCutoff	qvalue cutoff
readable	whether mapping gene ID to gene Name
use_internal_data	logical, if TRUE, use KEGG.db. default is FALSE, will download 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)
```

enrichMap

enrichMap

Description

enrichment map

Usage

```
enrichMap(x, n = 50, fixed = TRUE, ...)
```

Arguments

x	gseaResult or enrichResult object
n	maximum number of category to shown
fixed	if set to FALSE, will invoke tkplot
...	additional parameter

Details

enrichMap

Value

figure

Author(s)

ygc

getGOLevel	<i>get GOIDs at a specific level</i>
------------	--------------------------------------

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	<i>Gff2GeneTable</i>
---------------	----------------------

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

groupGO	<i>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.</i>
---------	--

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, organism = "human", ont = "CC", level = 2,
        readable = FALSE)
```

Arguments

gene	a vector of entrez gene id.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coelicolor", "ecolik12", "ecsakai", "fly", "gondii", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast" and "zebrafish".
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]], organism="human", 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,
     pvalueCutoff = 0.05, pAdjustMethod = "BH", TERM2GENE, TERM2NAME = NA,
     verbose = TRUE)
```

Arguments

geneList	order ranked geneList
exponent	weight of each step
nPerm	number of permutations
minGSSize	minimal size of each geneSet for analyzing
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

Value

gseaResult object

Author(s)

Guangchuang Yu

gseaplot

gseaplot

Description

visualize analyzing result of GSEA

Usage

```
gseaplot(gseaResult, geneSetID, by = "all")
```

Arguments

gseaResult	gseaResult object
geneSetID	geneSet ID
by	one of "runningScore" or "position"

Details

plotting function for gseaResult

Value

ggplot2 object

Author(s)

ygc

gseGO

*gseGO***Description**

Gene Set Enrichment Analysis of Gene Ontology

Usage

```
gseGO(geneList, ont = "BP", organism = "human", exponent = 1,
      nPerm = 1000, minGSSize = 10, pvalueCutoff = 0.05,
      pAdjustMethod = "BH", verbose = TRUE)
```

Arguments

geneList	order ranked geneList
ont	one of "BP", "MF", "CC" or "GO"
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coelicolor", "ecolik12", "ecsakai", "fly", "gondii", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast" and "zebrafish".
exponent	weight of each step
nPerm	permutation numbers
minGSSize	minimal size of each geneSet for analyzing
pvalueCutoff	pvalue Cutoff
pAdjustMethod	pvalue adjustment method
verbose	print message or not

Value

gseaResult object

Author(s)

Yu Guangchuang

gseKEGG

gseKEGG

Description

Gene Set Enrichment Analysis of KEGG

Usage

```
gseKEGG(geneList, organism = "human", exponent = 1, nPerm = 1000,  
        minGSSize = 10, pvalueCutoff = 0.05, pAdjustMethod = "BH",  
        use_internal_data = FALSE, verbose = TRUE)
```

Arguments

geneList	order ranked geneList
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "ecoli12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast" and "zebrafish".
exponent	weight of each step
nPerm	permutation numbers
minGSSize	minimal size of each geneSet for analyzing
pvalueCutoff	pvalue Cutoff
pAdjustMethod	pvalue adjustment method
use_internal_data	whether use KEGG.db or not
verbose	print message or not

Value

gseaResult object

Author(s)

Yu Guangchuang

idType	<i>idType</i>
--------	---------------

Description

list ID types supported by annoDb

Usage

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

Arguments

annoDb	annotation db
--------	---------------

Value

character vector

Author(s)

Guangchuang Yu

plot	<i>plot method</i>
------	--------------------

Description

plot method generics

Usage

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

Arguments

x	compareClusterResult object
type	one of bar or dot
title	figure title
font.size	font size
showCategory	category numbers

by	one of geneRatio, Percentage or count
colorBy	one of pvalue or p.adjust
includeAll	logical
...	Additional argument list

Value

plot

Author(s)

Guangchuang Yu <http://ygc.name>

plotting.clusterProfile

plotting-clusterProfile

Description

Internal plot function for plotting compareClusterResult

Usage

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

Arguments

clProf.reshape.df	data frame of compareCluster result
type	one of dot and bar
by	one of percentage and count
colorBy	one of pvalue or p.adjust
title	graph title
font.size	graph font size

Value

ggplot object

Author(s)

Guangchuang Yu <http://ygc.name>

viewKEGG	<i>viewKEGG function is for visualize KEGG pathways works with enrichResult object to visualize enriched KEGG pathway</i>
----------	---

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

- *Topic **classes**
 - compareClusterResult-class, [7](#)
 - groupGOResult-class, [15](#)
- *Topic **datasets**
 - DataSet, [7](#)
- *Topic **manip**
 - compareCluster, [6](#)
 - enrichGO, [10](#)
 - enrichKEGG, [11](#)
 - groupGO, [14](#)
- *Topic **package**
 - clusterProfiler-package, [2](#)

- bitr, [3](#)
- buildGOMap, [4](#)
- buildKEGGmap, [4](#)

- clusterProfiler
 - (clusterProfiler-package), [2](#)
- clusterProfiler-package, [2](#)
- cnetplot, [5](#)
- compareCluster, [6](#), [7](#), [10](#), [12](#), [14](#), [15](#)
- compareClusterResult, [3](#), [15](#)
- compareClusterResult-class, [7](#)

- DataSet, [7](#)
- download.KEGG, [8](#)

- enrichDAVID, [8](#)
- enricher, [9](#)
- enrichGO, [6](#), [10](#)
- enrichKEGG, [11](#)
- enrichMap, [12](#)
- enrichResult, [3](#), [7](#)

- gcSample (DataSet), [7](#)
- getGOLevel, [13](#)
- Gff2GeneTable, [13](#)
- groupGO, [6](#), [14](#), [15](#)
- groupGOResult, [3](#), [7](#)
- groupGOResult-class, [15](#)

- GSEA, [15](#)
- gseaplot, [16](#)
- gseGO, [17](#)
- gseKEGG, [18](#)

- idType, [19](#)

- plot, [19](#)
- plot, compareClusterResult, ANY-method (plot), [19](#)
- plot, compareClusterResult-method (compareClusterResult-class), [7](#)
- plotting.clusterProfile, [20](#)

- show, compareClusterResult-method (compareClusterResult-class), [7](#)
- show, groupGOResult-method (groupGOResult-class), [15](#)
- summary, compareClusterResult-method (compareClusterResult-class), [7](#)

- viewKEGG, [21](#)