

# Package ‘genekitr’

April 22, 2022

**Type** Package

**Title** Gene Analysis Toolkit in R

**Version** 0.6.1

**Maintainer** Yunze Liu <jieandze1314@gmail.com>

**URL** <https://github.com/GangLiLab/genekitr>

**BugReports** <https://github.com/GangLiLab/genekitr/issues>

**Description** An analysis toolkit based on the gene. It mainly includes five features (search, transform, analysis, visualization, and export). The user just needs to provide gene id ('entrez', 'symbol' or 'ensembl') then can search gene-related information, transform id type, do enrichment analysis, visualize enrichment result, visualize multi-group overlap, and export result as excel sheets to easily share and communicate with others.

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**Depends** R (>= 3.5)

**Imports** clusterProfiler, dplyr, ggplot2, stringr, stringi, tidyr, tibble, VennDiagram, rlang

**Suggests** BiocManager, cowplot, DOSE, data.table, easyPubMed, fgsea, futile.logger, ggplotify, ggsci, ggupset, ggrepel, magrittr, msigdbr, openxlsx, RColorBrewer, rappdirs, rentrez, scales, stats, utils, testthat (>= 3.0.0), knitr, rmarkdown

**RoxygenNote** 7.1.2

**NeedsCompilation** no

**Author** Yunze Liu [aut, cre]

**Repository** CRAN

**Date/Publication** 2022-04-22 04:10:02 UTC

## R topics documented:

as.enrichdat . . . . .	2
Datasets . . . . .	3
expoSheet . . . . .	3
genGO . . . . .	4
genGSEA . . . . .	5
genInfo . . . . .	6
genKEGG . . . . .	7
genPubmed . . . . .	9
plotEnrich . . . . .	9
plotGSEA . . . . .	11
plotVenn . . . . .	12
plot_theme . . . . .	13
transId . . . . .	14
<b>Index</b>	<b>16</b>

---

as.enrichdat	<i>Adjust dataframe for enrichment plot</i>
--------------	---

---

### Description

make sure colname contains Description, Count, FoldEnrich/GeneRatio, pvalue/qvalue/p.adjust

### Usage

```
as.enrichdat(enrich_df)
```

### Arguments

enrich\_df      dataframe of enrichment analysis result .

### Value

A 'data.frame'.

---

Datasets	<i>Datasets geneList entrez gene list with decreasing fold change value</i>
----------	---

---

**Description**

Datasets geneList entrez gene list with decreasing fold change value

Datasets msig\_species contains msigdb species information

Datasets msig\_category contains msigdb category information

Datasets biocOrg\_name contains organism name of bioconductor

Datasets keggOrg\_name contains organism name of KEGG [https://www.genome.jp/kegg/catalog/org\\_list.html](https://www.genome.jp/kegg/catalog/org_list.html)

Datasets ensOrg\_name contains organism name of ensembl

---

expoSheet	<i>Export list of datasets into different Excel sheets</i>
-----------	--

---

**Description**

Export list of datasets into different Excel sheets

**Usage**

```
expoSheet(
  data_list,
  name_list,
  filename = NULL,
  dir = tempdir(),
  overwrite = TRUE
)
```

**Arguments**

data_list	List of datasets.
name_list	List of data names.
filename	A character string naming an xlsx file.
dir	A character string naming output directory.
overwrite	If TRUE, overwrite any existing file.

**Value**

An Excel file.

**Examples**

```
## Not run:
library(openxlsx)
expoSheet(
  data_list = list(mtcars, ToothGrowth),
  name_list = list("mtcars", "tooth"),
  filename = "test.xlsx", dir = tempdir()
)

## End(Not run)
```

genGO

*Gene GO enrichment analysis***Description**

Gene GO enrichment analysis

**Usage**

```
genGO(
  id,
  group_list = NULL,
  org,
  ont,
  use_symbol = TRUE,
  pAdjustMethod = "BH",
  pvalueCutoff = 0.05,
  qvalueCutoff = 0.1,
  minGSSize = 10,
  maxGSSize = 500,
  universe,
  ...
)
```

**Arguments**

id	A vector of gene id which can be entrez, ensembl or symbol.
group_list	A list of gene id groups, default is NULL.
org	Organism name from 'biocOrg_name'.
ont	One of "bp", "mf", and "cc" subontologies, or "all" for all three.
use_symbol	Logical to set result gene id as gene symbol, default is TRUE.
pAdjustMethod	One of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none".
pvalueCutoff	Adjusted pvalue cutoff, default is 0.05.
qvalueCutoff	Adjusted pvalue cutoff, default is 0.1.

minGSSize	Minimal size of each gene set for analyzing, default is 10.
maxGSSize	Maximal size of each gene set for analyzing, default is 500.
universe	Background genes. If missing, then all gene list in orgdb will be used as background.
...	other argument to 'enrichGO' function

### Value

A 'data.frame' contains gene ratio and fold enrichment.

### Examples

```
## Not run:
data(geneList, package = "genekitr")

# only gene ids
id <- names(geneList)[1:100]
ego <- genGO(id,
  org = "human", ont = "cc", pvalueCutoff = 0.01,
  qvalueCutoff = 0.1, use_symbol = FALSE
)
head(ego)

# gene id with groups
id <- c(head(names(geneList),50),tail(names(geneList),50))
group <- list(group1 = c(rep('up',50),rep('down',50)),
  group2 = c(rep('A',40),rep('B',60)))

gego <- genGO(id, group_list = group,
  org = "human", ont = "bp", pvalueCutoff = 0.1,
  qvalueCutoff = 1, use_symbol = FALSE
)

## End(Not run)
```

---

genGSEA

*GSEA for a gene list with decreasing logFC value*

---

### Description

GSEA for a gene list with decreasing logFC value

### Usage

```
genGSEA(
  genelist,
  org,
```

```

category = c("C1", "C2", "C3", "C4", "C5", "C6", "C7", "C8", "H"),
subcategory = NULL,
use_symbol = TRUE,
minGSSize = 10,
maxGSSize = 500,
pvalueCutoff = 0.05,
...
)

```

### Arguments

genelist	Order ranked genelist in decreasing order, gene can be entrez, ensembl or symbol.
org	Organism name from 'msig_org'.
category	MSigDB collection abbreviation, one of 'C1','C2','C3','C4','C5','C6','C7','C8','H'.
subcategory	MSigDB sub-collection abbreviation, choose from 'msig_category'.
use_symbol	Logical to set result gene id as gene symbol, default is TRUE.
minGSSize	Minimal size of each geneSet for analyzing, default is 10.
maxGSSize	Maximal size of each geneSet for analyzing, default is 500.
pvalueCutoff	Adjusted pvalue cutoff, default is 0.05.
...	Other argument to 'GSEA' function

### Value

GSEA list

### Examples

```

data(geneList, package = "genekitr")
gse <- genGSEA(genelist = geneList, org = "human",
  category = "H",use_symbol = TRUE)

```

---

genInfo

*Get gene related information*

---

### Description

Get gene related information

### Usage

```
genInfo(id = NULL, org, unique = FALSE)
```

**Arguments**

id	Gene id (symbol, ensembl or entrez id) or uniprot id. If this argument is NULL, return all gene info.
org	Short latin name from 'ensOrg_name_data'.
unique	Logical to keep only one matched ID, default is FALSE.

**Value**

A 'data.frame'.

**Examples**

```
# input id contains fake id and one-to-many match id
x <- genInfo(id = c(
  "MCM10", "CDC20", "S100A9", "MMP1", "BCC7",
  "FAKEID", "TP53", "HBD", "NUDT10"
), org = "hg", unique = TRUE)
head(x)
```

---

genKEGG

*Gene enrichment of KEGG analysis*

---

**Description**

Gene enrichment of KEGG analysis

**Usage**

```
genKEGG(
  id,
  group_list = NULL,
  org,
  use_symbol = TRUE,
  pAdjustMethod = "BH",
  pvalueCutoff = 0.05,
  qvalueCutoff = 0.1,
  minGSSize = 10,
  maxGSSize = 500,
  universe,
  ...
)
```

**Arguments**

id	A vector of entrez gene.
group_list	A list of gene id groups, default is NULL.
org	KEGG organism name from 'keggOrg_name'.
use_symbol	Logical to set result gene id as gene symbol, default is TRUE.
pAdjustMethod	One of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none".
pvalueCutoff	Numeric of adjusted pvalue cutoff, default is 0.05.
qvalueCutoff	Numeric of adjusted pvalue cutoff, default is 0.1.
minGSSize	Numeric of minimal size of each geneSet for analyzing, default is 10.
maxGSSize	Numeric of maximal size of each geneSet for analyzing, default is 500.
universe	Background genes. If missing, the orgdb all gene list will be used as background.
...	Other argument to 'enrichKEGG' function

**Value**

A 'data.frame'.

**Examples**

```
## Not run:
# only gene ids
data(geneList, package = "genekitr")
id <- names(geneList)[1:100]
keg <- genKEGG(id, org = "human")

# gene id with groups
id <- c(head(names(geneList),50),tail(names(geneList),50))
group <- list(group1 = c(rep('up',50),rep('down',50)),
              group2 = c(rep('A',40),rep('B',60)))
gkeg <- genKEGG(id, group_list = group,
                org = "human", pvalueCutoff = 0.1,
                qvalueCutoff = 0.1, use_symbol = FALSE
)

## End(Not run)
```



---

genPubmed	<i>Gene related pubmed paper records</i>
-----------	--

---

**Description**

Gene related pubmed paper records

**Usage**

```
genPubmed(id, keywords, field = "ALL")
```

**Arguments**

id	Gene "symbol".
keywords	species "mm" or "hs".
field	pubmed field from 'showNCBI('pubmed')', default is "ALL".

**Value**

A 'data.frame'.

**Examples**

```
x <- genPubmed(  
  id = c("Cyp2c23", "Fhit", "Gal3st2b", "Insl3", "Gbp4"),  
  keywords = "stem cell", field = "tiab"  
)
```

---

plotEnrich	<i>Dotplot for GO and KEGG enrichment analysis</i>
------------	--

---

**Description**

Dotplot for GO and KEGG enrichment analysis

**Usage**

```
plotEnrich(  
  enrich_df,  
  plot_type = c("bar", "dot"),  
  xlab_type = c("FoldEnrich", "GeneRatio", "Count"),  
  legend_type = c("p.adjust", "pvalue", "qvalue"),  
  low_color = "red",  
  high_color = "blue",
```

```

    show_item = 10,
    xlim_left = 0,
    xlim_right = NA,
    wrap_length = NULL,
    ...
  )

```

### Arguments

enrich_df	‘data.frame’ of enrichment analysis result.
plot_type	One of "bar" and "dot"
xlab_type	X-axis label type, one of 'GeneRatio', 'Count', 'FoldEnrich'.
legend_type	Stats legend type, one of "pvalue", "p.adjust", "qvalue".
low_color	Legend color for low pvalue or qvalue, default is "red".
high_color	Legend color for high pvalue or qvalue, default is "blue".
show_item	Numeric, select top N rows to show, default is 10.
xlim_left	X-axis left limit, default is 0.
xlim_right	X-axis right limit, default is NA.
wrap_length	Numeric, wrap text if longer than this length, default is NULL.
...	other arguments transfer to ‘plot_theme’ function

### Value

A ggplot object

### Examples

```

## Not run:
library(ggplot2)
data(geneList, package = "genekitr")
id <- names(geneList)[1:100]
ego <- genGO(id,
  org = "human", ont = "bp", pvalueCutoff = 0.05,
  qvalueCutoff = 0.05, use_symbol = TRUE
)
plotEnrich(ego, plot_type = "dot")

plotEnrich(ego, plot_type = "bar")

## End(Not run)

```

---

plotGSEA

*GSEA plot*

---

## Description

GSEA plot

## Usage

```
plotGSEA(  
  gsea_list,  
  plot_type = c("volcano", "classic", "fgsea"),  
  show_pathway = 3,  
  show_genes = NULL,  
  colors = NULL,  
  ...  
)
```

## Arguments

<code>gsea_list</code>	GSEA result from 'genGSEA' function
<code>plot_type</code>	GSEA plot type, one of 'volcano', 'classic' or 'fgsea'.
<code>show_pathway</code>	Which pathways included, user could specify number (default is 3) or character name.
<code>show_genes</code>	Character to specify gene names included in plot when 'plot_type' is "pathway".
<code>colors</code>	Character to specify colors when 'plot_type' is "pathway".
<code>...</code>	other arguments transfer to 'plot_theme' function

## Value

A ggplot object

## Examples

```
## Not run:  
library(ggplot2)  
# get GSEA result  
data(geneList, package = "genekitr")  
gse <- genGSEA(genelist = geneList, org = "human",  
              category = "H", use_symbol = TRUE, pvalueCutoff = 1)  
# volcano plot  
plotGSEA(gse, plot_type = c('volcano'), show_pathway = 3)  
  
# classic pathway plot  
plotGSEA(gse, plot_type = c('classic'), show_pathway = 1:2)  
  
# fgsea for multiple pathway
```

```
plotGSEA(gse, plot_type = c('fgsea'), show_pathway = 10)

## End(Not run)
```

---

plotVenn

*Venn plot for groups of genes*


---

### Description

If gene group over 4, plot will be visualized using UpSet plot.

### Usage

```
plotVenn(
  venn_list,
  use_venn = TRUE,
  color = NULL,
  alpha_degree = 0.3,
  text_size = 1,
  border_thick = 1,
  remove_grid = FALSE,
  ...
)
```

### Arguments

venn_list	A list of gene id.
use_venn	Logical, use venn to plot, default is 'TRUE', the other option is upsetplot for large list.
color	Colors for gene lists, default is NULL.
alpha_degree	Alpha transparency of each circle's area, default is 0.3.
text_size	Text size, default is 1.
border_thick	Numeric, border thickness, default is 1.
remove_grid	Logical, remove circle or grid lines, default is 'FALSE'.
...	other arguments transfer to 'plot_theme' function

### Value

A ggplot object

**Examples**

```

library(ggplot2)
set1 <- paste0(rep("gene", 100), sample(c(1:1000), 100))
set2 <- paste0(rep("gene", 100), sample(c(1:1000), 100))
set3 <- paste0(rep("gene", 100), sample(c(1:1000), 100))
set4 <- paste0(rep("gene", 100), sample(c(1:1000), 100))
set5 <- paste0(rep("gene", 100), sample(c(1:1000), 100))
sm_gene_list <- list(gset1 = set1, gset2 = set2, gset3 = set3)
la_gene_list <- list(
  gset1 = set1, gset2 = set2, gset3 = set3,
  gset4 = set4, gset5 = set5
)
plotVenn(sm_gene_list,
  text_size = 1.5, alpha_degree = 1,
  remove_grid = TRUE, color = ggsci::pal_lancet()(3)
)
plotVenn(la_gene_list,
  text_size = 15, alpha_degree = 0.2, border_thick = 2,
  remove_grid = TRUE, use_venn = FALSE
)

```

---

plot\_theme

*Themes for all plots*


---

**Description**

Change ggplot text, font, legend and border

**Usage**

```

plot_theme(
  theme_type = c("bw", "classic"),
  main_text_size = 8,
  legend_text_size = 6,
  font_type = "sans",
  border_thick = 1.5,
  remove_grid = TRUE,
  remove_border = FALSE,
  remove_text = FALSE,
  remove_legend = FALSE
)

```

**Arguments**

theme\_type      ggplot theme,  
main\_text\_size   Numeric, main text size  
legend\_text\_size      Numeric, legend text size

font_type	Character, specify the plot text font family, default is "sans".
border_thick	Numeric, border thickness, default is 1.
remove_grid	Logical, remove background grid lines, default is FALSE.
remove_border	Logical, remove border line, default is FALSE.
remove_text	Logical, remove all text, default is FALSE.
remove_legend	Logical, remove legend, default is FALSE.

**Value**

ggplot theme

**Examples**

```
library(ggplot2)
ggplot(mtcars, aes(x=wt, y=mpg))+ geom_point()+
  plot_theme(theme_type = 'bw', font_type = 'Times', border_thick = 2)
```

---

transId

---

*Transform gene id among symbol, entrezid, ensembl and uniprot*


---

**Description**

Transform gene id among symbol, entrezid, ensembl and uniprot

**Usage**

```
transId(id, trans_to, org, unique = TRUE)
```

**Arguments**

id	Gene ids.
trans_to	Transform to which type, one of "symbol", "entrezid", "ensembl" and "uniprot."
org	Short latin name from 'ensOrg_name_data'.
unique	If keep only one unique mapped ID when one-to-many gene occurs, default is FALSE.

**Value**

A character of transformed ids.

**Examples**

```
## Not run:
transId(
  id = c("Cyp2c23", "Fhit", "Gal3st2b", "Trp53", "Tp53"),
  trans_to = "ensembl", org = "mouse", unique = TRUE
)
# input id contains fake id and one-to-many match id
transId(
  id = c("MMD2", "HBD", "RNR1", "TEC", "BCC7", "FAKEID", "TP53"),
  trans_to = "entrez", org = "hg", unique = FALSE
)
# recognize ensembl version number
transId('ENSG00000141510.11', 'symbol', 'hs')

## End(Not run)
```

# Index

## \* datasets

Datasets, [3](#)

as.enrichdat, [2](#)

biocOrg\_name (Datasets), [3](#)

Datasets, [3](#)

ensOrg\_name (Datasets), [3](#)

expoSheet, [3](#)

geneList (Datasets), [3](#)

genGO, [4](#)

genGSEA, [5](#)

genInfo, [6](#)

genKEGG, [7](#)

genPubmed, [9](#)

keggOrg\_name (Datasets), [3](#)

msig\_category (Datasets), [3](#)

msig\_org (Datasets), [3](#)

plot\_theme, [13](#)

plotEnrich, [9](#)

plotGSEA, [11](#)

plotVenn, [12](#)

transId, [14](#)