

Package ‘DSEATM’

July 28, 2022

Title drug set enrichment analysis based on text mining

Version 1.0.0

Depends R (>= 3.4.0),clusterProfiler,RSQLite,AnnotationDbi,org.Hs.eg.db,ensurer

Description We have extract the disease-drug and drug-gene associations from MEDLINE publications. the DSEATM has been implemented to display and use these data.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

Suggests rmarkdown,
knitr

VignetteBuilder knitr

R topics documented:

cmap_drug	2
disease2drug	2
disease2gene	3
disease2pathway	4
drug2gene	5
loadDisease2drugDb	5
pathway_gene_df	6
Index	7

cmap_drug	<i>the CMap drug MesH ID</i>
-----------	------------------------------

Description

A dataset containing the cmap drug name and MeSH term ID there are 1309 drugs in CMap, 1129 of them can be assigned MeSH ID cmap_drug

Usage

```
cmap_drug
```

Format

A data frame with 1129 rows and 4 variables:

cmap_name drug names, this is the drug original names

CID drug pubchem ID

meshterm drug MeSH name, parsed by pyMeSHSim

meshID drug MeSH ID, the drug MeSH ID ...

disease2drug	<i>Select the disease related drugs</i>
--------------	---

Description

Select the disease related drugs

Usage

```
disease2drug(DiseaseMeSHID, Repository = "DSEATM")
```

Arguments

DiseaseMeSHID The MeSH id of intersted disease, only Main heading diseases are allowed here.

Repository one of the values in c("ALL", "DSEATM"). "ALL" means all MeSH terms in "D" category will be obtained in the disease-drug association. In this condition, all the terms are chemical and drugs. "DSEATM" means only MeSH terms below "ChemicalActionsandUses", "OrganicChemicals" and "Compounds,Heterocyclic" The default is "DSEATM"

Value

dataframe It contained all drugs associated to a specific diseases.

Examples

```
#here, we only give out the drug ID in the table,
#detail information for each drug can be retrived by pyMeSHSim
#get disease-drug association in DSEATM repository
df = disease2drug("D015179")
# when using ALL repository,
# the disease-drug associations are in indeed disease-chemical associations
# Users can filter the drugs by their own principles in this data set.
df = disease2drug("D015179", Repository="ALL")

#get CMap drugs associated to a disease
df = disease2drug("D015179", Repository="ALL")
df_cmap = df[df$Drug %in% DSEATM::cmap_drug$meshID,]
```

disease2gene	<i>Select the disease related genes by drugs</i>
--------------	--

Description

Select the disease related genes by drugs

Usage

```
disease2gene(DiseaseMeSHID, Repository = "DSEATM")
```

Arguments

DiseaseMeSHID	the MeSH id of intersted disease
Repository	one of the values in c("ALL", "DSEATM"). "ALL" means all MeSH terms in "D" category will be obtained in the disease-drug association. In this condition, all the terms are chemical and drugs. "DSEATM" means only MeSH terms below "ChemicalActionsandUses", "OrganicChemicals" and "Compounds,Heterocyclic" The default is "DSEATM"

Value

dataframe It containing MeSHID, GENEID and group. MESHID is MeSH drug and chemical terms ID. GENEID is Entrez gene ID. group is the drug-gene relation source. fisher_test means relations extracted from publication abstracts, meshr means relations extracted from MeSH.Hsg.eg.db

Examples

```
disease2gene("D015179")
```

disease2pathway

*Enriched the pathways for diseases***Description**

The disease2pathway function is used to calculate the enrichment of disease genes. Given an input gene set, it invokes the clusterProfiler to realize the enrichment function. The gene set can be retrieved by a disease name. DSEATM used a diseases associated drugs to get associated genes,

Usage

```
disease2pathway(
  DiseaseMeSHID,
  Repository = "DSEATM",
  data_source = "ALL",
  gene_freq = 1
)
```

Arguments

DiseaseMeSHID	the MeSH id of interested disease, only MeSH main headings are allowed here.
Repository	one of the values in c("ALL", "DSEATM"). "ALL" means all MeSH terms in "D" category will be obtained in the disease-drug association. In this condition, all the terms are chemical and drugs. "DSEATM" means only MeSH terms below "ChemicalActionsandUses", "OrganicChemicals" and "Compounds,Heterocyclic". The default is "DSEATM"
data_source	It denotes the drug-gene relation source used in enrichment, one of the values in c("ALL", "mineR", "meshr") "ALL" means all drug-gene relations are used to extract drug-gene association. "mineR" means only drug-gene relations extracted by pubmed.mineR from publication abstracts were used. "meshr" means only drug-gene relations extracted from MeSH.Hsg.eg.db were used. The default is "ALL"
gene_freq	An integer describing the frequency threshold of genes used in enrichment. A higher gene_freq means genes associated with more MeSH drug and chemical terms are used. The default is 1

Value

dataframe It containing the pathways enriched

Author(s)

Zhi-Hui Luo, Ya-Min Wang

Examples

```
# get pathways enriched for breast cancer
disease2pathway("D001943")
disease2pathway("D001943", data_source="ALL", gene_freq=2)
```

drug2gene	Select the drug related genes
-----------	-------------------------------

Description

Select the drug related genes

Usage

```
drug2gene(DrugMeSHID)
```

Arguments

DrugMeSHID the MeSH id of interested drug

Value

dataframe

Examples

```
drug2gene("D016861")
```

loadDisease2drugDb	This is for load loadDisease2drugDb
--------------------	-------------------------------------

Description

The disease-drug and drug-gene associations are the core data of DSEATM. We have generated two kind of relations for disease-drug, and drug-gene associations. The one is "ALL" data, the other is "DSEATM" data.

"ALL" means all MeSH terms in D category were used. "DSEATM" means only part of MeSH terms in "ChemicalActionsandUses", "OrganicChemicals" and "Compounds,Heterocyclic" were used.

Usage

```
loadDisease2drugDb(data_type, Repository = "DSEATM")
```

Arguments

data_type	One of the values in c("Disease", "Drug"). This parameter is used to get disease2drug table or drug2gene table. table disease2drug is in Disease type database. Five columns in disease2drug: Drug, Disease, CoFre, oddratio, pvalue Disease is the MeSH disease term ID Drug is the MeSH Drug term ID CoFree is the co-occurrence of one disease-drug pair, oddratio is the fisher test of the co-occurrence pvalue is the significance of the fisher test table drug2gene is in Drug type database. Three columns in disease2drug: MESHID, GENEID, group MESHID is the MeSH drugs, GENEID is the Entrez gene ID group is the source of the drug-gene association all drug-gene association in this table are significant relations.
Repository	one of the values in c("ALL", "DSEATM"). "ALL" means all MeSH terms in "D" category will be obtained in the disease-drug association. In this condition, all the terms are chemical and drugs. "DSEATM" means only MeSH terms below "ChemicalActionsandUses", "OrganicChemicals" and "Compounds,Heterocyclic" The default is "DSEATM"

Examples

```
dis <- "D015179"
con <- loadDisease2drugDb(data_type="Disease")
res <- RSQLite::dbSendQuery(con,
  sprintf("SELECT * FROM disease2drug WHERE Disease = %s%s%s", "'", dis, "'"))
df = as.data.frame(RSQLite::dbFetch(res))
```

pathway_gene_df	<i>the genes included in each pathway</i>
-----------------	---

Description

A dataset containing the genes and pathways pathway_gene_df

Usage

```
pathway_gene_df
```

Format

A data frame with 506182 rows and 2 variables:

pathway_name pathway names, MSigDB C5 GO biological process

gene one gene in the pathway, gene symbol ...

Source

<https://www.gsea-msigdb.org/gsea/msigdb/>

Index

* **datasets**

cmap_drug, [2](#)

pathway_gene_df, [6](#)

cmap_drug, [2](#)

disease2drug, [2](#)

disease2gene, [3](#)

disease2pathway, [4](#)

drug2gene, [5](#)

loadDisease2drugDb, [5](#)

pathway_gene_df, [6](#)