# Package 'geneset'

July 27, 2026

**Type** Package

**Title** Get Gene Sets for Gene Enrichment Analysis

**Version** 0.2.8

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

**URL** https://github.com/GangLiLab/geneset

**BugReports** https://github.com/GangLiLab/geneset/issues

**Description**

Gene sets are fundamental for gene enrichment analysis. The package 'geneset' enables querying gene sets from public databases including 'GO' (Gene Ontology Consortium. (2004) <doi:10.1093/nar/gkh036>),
'KEGG' (Minoru et al. (2000) <doi:10.1093/nar/28.1.27>),
'WikiPathway' (Marvin et al. (2020) <doi:10.1093/nar/gkaa1024>),
'MsigDb' (Arthur et al. (2015) <doi:10.1016/j.cels.2015.12.004>),
'Reactome' (David et al. (2011) <doi:10.1093/nar/gkq1018>),
'MeSH' (Ish et al. (2014) <doi:10.4103/0019-5413.139827>),
'DisGeNET' (Janet et al. (2017) <doi:10.1093/nar/gkw943>),
'Disease Ontology' (Lynn et al. (2011) <doi:10.1093/nar/gkr972>),
'Network of Cancer Genes' (Dimitra et al. (2019) <doi:10.1186/s13059-018-1612-0>) and
'COVID-19' (Maxim et al. (2020) <doi:10.21203/rs.3.rs-28582/v1>).
Gene sets are stored in the list object which provides data frame of 'geneset' and 'geneset_name'. The 'geneset' has two columns of term ID and gene ID. The 'geneset_name' has two columns of terms ID and term description.

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**Depends** R (>= 3.6)

**Imports** dplyr, RCurl, fst, stringr

**Suggests** testthat (>= 3.0.0)

**Config/testthat/edition** 3

**Config/roxygen2/version** 8.0.0

**NeedsCompilation** no

**Author** Yunze Liu [aut, cre] (ORCID: <https://orcid.org/0000-0002-7414-8556>)

**Repository** CRAN

**Date/Publication** 2026-07-27 09:50:09 UTC

# Contents

Datasets . . . . . 2
getEnrichrdb . . . . . 3
getGO . . . . . 3
getHgDisease . . . . . 4
getKEGG . . . . . 5
getMesh . . . . . 6
getMsigdb . . . . . 7
getReactome . . . . . 8
getWiki . . . . . 9

**Index** **10**

---

`Datasets` *Datasets go_org contains GO species information*

---

## Description

Datasets go_org contains GO species information

Datasets kegg_org contains KEGG species information

Datasets wiki_org contains WikiPathway species information

Datasets msigdb_org contains Msigdb species information

Datasets enrichr_metadata contains Enrichrdb information

Datasets reactome_org contains Reactome species information

Datasets mesh_org contains MeSH species information

Datasets mesh_metadata contains MeSH information

Datasets org2cate contains all organism to category information

Datasets org2cate contains all organism to category information

Datasets ensOrg_name contains organism name of ensembl

`getEnrichrdb` *Get EnrichrDB geneset and geneset_name Geneset is a data.frame of 2 columns with term id and gene id*

## Description

Get EnrichrDB geneset and geneset_name Geneset is a data.frame of 2 columns with term id and gene id

## Usage

```
getEnrichrdb(
  org = c("human", "fly", "yeast", "worm", "zebrafish"),
  library = NULL,
  download.method = NULL,
  data_dir = NULL
)
```

## Arguments

| | |
|---|---|
| `org` | Organism from 'human','fly','yeast','worm','zebrafish'. |
| `library` | Choose one library name from 'enrichr_metadata'. |
| `download.method` | "auto" (as default if NULL), "wininet" (for windows) |
| `data_dir` | data saving location, default is the package data directory |

## Value

A list including geneset and geneset name.

## Examples

```
x = getEnrichrdb(org = "human", library = "COVID-19_Related_Gene_Sets",
data_dir = tempdir())
```

`getGO` *Get GO geneset and geneset_name Geneset is a data.frame of 2 columns with term id and gene id; Geneset_name is a data.frame of 2 columns with term id and term description*

## Description

Get GO geneset and geneset_name Geneset is a data.frame of 2 columns with term id and gene id; Geneset_name is a data.frame of 2 columns with term id and term description

## Usage

```
getGO(
  org = "human",
  ont = c("bp", "cc", "mf"),
  download.method = NULL,
  data_dir = NULL
)
```

## Arguments

| | |
|---|---|
| org | Organism name from ‘go_org‘. |
| ont | One of "bp","cc" or "mf" ontology. |
| download.method | "auto" (as default if NULL), "wininet" (for windows) |
| data_dir | data saving location, default is the package data directory |

## Value

A list including geneset and geneset name.

## Examples

```
x = getGO(org = "human",ont = "mf", data_dir = tempdir())
```

---

# getHgDisease *Get HgDisease geneset and geneset_name Human disease gene sets from Disease Ontology (DO),DisGeNET, Network of Cancer Gene (NCG) version 6 and v7, and covid19-specific. Geneset is a data.frame of 2 columns with term id and gene id; Geneset_name is a data.frame of 2 columns with term id and term description*

---

## Description

Get HgDisease geneset and geneset_name Human disease gene sets from Disease Ontology (DO),DisGeNET, Network of Cancer Gene (NCG) version 6 and v7, and covid19-specific. Geneset is a data.frame of 2 columns with term id and gene id; Geneset_name is a data.frame of 2 columns with term id and term description

## Usage

```
getHgDisease(
  source = c("do", "disgenet", "ncg_v7", "ncg_v6", "covid19"),
  download.method = NULL,
  data_dir = NULL
)
```

### Arguments

| | |
|---|---|
| source | Choose from 'do','ncg_v7','ncg_v6','disgenet','covid19'. |
| download.method | "auto" (as default if NULL), "wininet" (for windows) |
| data_dir | data saving location, default is the package data directory |

### Value

A list.

### Examples

```
x = getHgDisease(source = "do", data_dir = tempdir())
```

---

## getKEGG *Get KEGG geneset and geneset_name Geneset is a data.frame of 2 columns with term id and gene id; Geneset_name is a data.frame of 2 columns with term id and term description*

---

### Description

Get KEGG geneset and geneset_name Geneset is a data.frame of 2 columns with term id and gene id; Geneset_name is a data.frame of 2 columns with term id and term description

### Usage

```
getKEGG(
  org = "hsa",
  category = c("pathway", "module", "enzyme", "disease", "drug", "network"),
  download.method = NULL,
  data_dir = NULL
)
```

### Arguments

| | |
|---|---|
| org | Organism name from 'kegg_org'. |
| category | Choose one category from "pathway","module", "enzyme", 'disease' (human only),'drug' (human only) or 'network' (human only) |
| download.method | "auto" (as default if NULL), "wininet" (for windows) |
| data_dir | data saving location, default is the package data directory |

### Value

A list including geneset and geneset name.

## Examples

```
x = getKEGG(org = "hsa",category = "pathway", data_dir = tempdir())
```

---

getMesh *Get MeSH geneset and geneset_name MeSH is the annotation used for MEDLINE/PubMed articles and is manually curated by NLM (U.S. National Library of Medicine). Geneset is a data.frame of 2 columns with term id and gene id; Geneset_name is a data.frame of 2 columns with term id and term description*

---

## Description

Get MeSH geneset and geneset_name MeSH is the annotation used for MEDLINE/PubMed articles and is manually curated by NLM (U.S. National Library of Medicine). Geneset is a data.frame of 2 columns with term id and gene id; Geneset_name is a data.frame of 2 columns with term id and term description

## Usage

```
getMesh(
  org = "human",
  method = c("gendoo", "gene2pubmed", "RBBH"),
 category = c("A", "B", "C", "D", "E", "F", "G", "H", "I", "J", "K", "L", "M", "N", "Z"),
  download.method = NULL,
  data_dir = NULL
)
```

## Arguments

| | |
|---|---|
| org | Organism mesh_name from ‘mesh_org‘. |
| method | Method of mapping MeSH ID to gene ID. Choose one from "gendoo", "gene2pubmed" or "RBBH" (mainly for some minor species). See also ‘mesh_metadata‘. |
| category | MeSH descriptor categories from ‘mesh_metadata‘ (refer to: https://wikipedia.org/wiki/List_of_MeSH_c |
| download.method | "auto" (as default if NULL), "wininet" (for windows) |
| data_dir | data saving location, default is the package data directory |

## Value

A list including geneset and geneset name.

## Examples

```
x = getMesh(org = "human", method = "gendoo", category = "A", data_dir = tempdir())
```

getMsigdb *Get MSigDB geneset and geneset_name*

## Description

Geneset is a data.frame of 2 columns with term id and gene id. The function keeps the original public interface and automatically distinguishes Human MSigDB collections from native Mouse MSigDB collections by ‘category‘.

## Usage

```
getMsigdb(
  org = "human",
  category = c("H", "C1", "C2-CGP", "C2-CP-BIOCARTA", "C2-CP-KEGG", "C2-CP-PID",
    "C2-CP-REACTOME", "C2-CP-WIKIPATHWAYS", "C3-MIR-MIRDB", "C3-MIR-MIR_Legacy",
   "C3-TFT-GTRD", "C3-TFT-TFT_Legacy", "C4-CGN", "C4-CM", "C5-GO-BP", "C5-GO-CC",
    "C5-GO-MF", "C5-HPO", "C6", "C7-IMMUNESIGDB", "C7-VAX", "C8"),
  download.method = NULL,
  data_dir = NULL
)
```

## Arguments

| | |
|---|---|
| org | Organism name from ‘msigdb_org‘. |
| category | MSigDB collection/subcollection name. Existing Human collection names remain supported; native Mouse collection names such as "MH", "M1", "M2-CP-REACTOME" and "M5-GO-BP" are also supported. |
| download.method | "auto" (as default if NULL), "wininet" (for windows) |
| data_dir | data saving location, default is the package data directory |

## Details

Human MSigDB collections (e.g. H, C1, C2-CP-REACTOME, C5-GO-BP) may contain ortholog-converted data for species supported by ‘msigdb_org‘.

Native Mouse MSigDB collections (e.g. MH, M1, M2-CP-REACTOME, M5-GO-BP) are only support Mus musculus.

## Value

A list including geneset.

## Examples

```
# Human native Hallmark
x <- getMsigdb(org = "human", category = "H", data_dir = tempdir())

# Human MSigDB converted to mouse orthologs
x <- getMsigdb(org = "mouse", category = "H", data_dir = tempdir())

# Native Mouse MSigDB
x <- getMsigdb(org = "mouse", category = "MH", data_dir = tempdir())
```

---

`getReactome` *Get Reactome geneset and geneset_name Geneset is a data.frame of 2 columns with term id and gene id; Geneset_name is a data.frame of 2 columns with term id and term description*

---

## Description

Get Reactome geneset and geneset_name Geneset is a data.frame of 2 columns with term id and gene id; Geneset_name is a data.frame of 2 columns with term id and term description

## Usage

```
getReactome(org = "human", download.method = NULL, data_dir = NULL)
```

## Arguments

| | |
|---|---|
| `org` | Organism name from ‘reactome_org‘. |
| `download.method` | "auto" (as default if NULL), "wininet" (for windows) |
| `data_dir` | data saving location, default is the package data directory |

## Value

A list including geneset and geneset name.

## Examples

```
x = getReactome(org = "human", data_dir = tempdir())
```

getWiki *Get WikiPathway geneset and geneset_name Geneset is a data.frame of 2 columns with term id and gene id; Geneset_name is a data.frame of 2 columns with term id and term description*

## Description

Get WikiPathway geneset and geneset_name Geneset is a data.frame of 2 columns with term id and gene id; Geneset_name is a data.frame of 2 columns with term id and term description

## Usage

```
getWiki(org = "human", download.method = NULL, data_dir = NULL)
```

## Arguments

| | |
|---|---|
| org | Organism name from 'wiki_org'. |
| download.method | "auto" (as default if NULL), "wininet" (for windows) |
| data_dir | data saving location, default is the package data directory |

## Value

A list including geneset and geneset name.

## Examples

```
x = getWiki(org = "human", data_dir = tempdir())
```

# Index

∗ **datasets**
Datasets, 2

Datasets, 2

enrichr_metadata *(*Datasets*)*, 2
ensOrg_name *(*Datasets*)*, 2

getEnrichrdb, 3
getGO, 3
getHgDisease, 4
getKEGG, 5
getMesh, 6
getMsigdb, 7
getReactome, 8
getWiki, 9
go_org *(*Datasets*)*, 2

kegg_org *(*Datasets*)*, 2

mesh_metadata *(*Datasets*)*, 2
mesh_org *(*Datasets*)*, 2
msigdb_org *(*Datasets*)*, 2

org2cate *(*Datasets*)*, 2

reactome_org *(*Datasets*)*, 2

wiki_org *(*Datasets*)*, 2