

Package ‘msigdbr’

April 15, 2025

Type Package

Title MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format

Version 10.0.2

Description Provides the 'Molecular Signatures Database' (MSigDB) gene sets typically used with the 'Gene Set Enrichment Analysis' (GSEA) software (Subramanian et al. 2005 <[doi:10.1073/pnas.0506580102](https://doi.org/10.1073/pnas.0506580102)>, Liberzon et al. 2015 <[doi:10.1016/j.cels.2015.12.004](https://doi.org/10.1016/j.cels.2015.12.004)>, Castanza et al. 2023 <[doi:10.1038/s41592-023-02014-7](https://doi.org/10.1038/s41592-023-02014-7)>) as an R data frame. The package includes the human genes as listed in MSigDB as well as the corresponding symbols and IDs for frequently studied model organisms such as mouse, rat, pig, fly, and yeast.

License MIT + file LICENSE

URL <https://igordot.github.io/msigdbr/>

BugReports <https://github.com/igordot/msigdbr/issues>

Depends R (>= 4.1)

Imports assertthat, babelgene (>= 22.9), dplyr (>= 1.1.1), lifecycle, methods, rlang, tibble, tidyselect (>= 1.2.0)

Suggests knitr, msigdbdf, rmarkdown, roxygen2, testthat

Additional_repositories <https://igordot.r-universe.dev>

Config/Needs/website rmarkdown

Encoding UTF-8

RoxygenNote 7.3.2

NeedsCompilation no

Author Igor Dolgalev [aut, cre] (<<https://orcid.org/0000-0003-4451-126X>>)

Maintainer Igor Dolgalev <igor.dolgalev@nyumc.org>

Repository CRAN

Date/Publication 2025-04-14 22:00:02 UTC

Contents

msigdb	2
msigdb_check_data	3
msigdb_collections	4
msigdb_species	4

Index	5
--------------	----------

msigdb	<i>Retrieve the gene sets data frame</i>
--------	--

Description

Retrieve a data frame of gene sets and their member genes. The original human genes can be converted into their corresponding counterparts in various model organisms, including mouse, rat, pig, zebrafish, fly, and yeast. The output includes gene symbols along with NCBI and Ensembl IDs.

Usage

```
msigdb(
  species = "Homo sapiens",
  db_species = "HS",
  collection = NULL,
  subcollection = NULL,
  category = deprecated(),
  subcategory = deprecated()
)
```

Arguments

species	Species name for output genes, such as "Homo sapiens" or "Mus musculus". Use <code>msigdb_species()</code> for available options.
db_species	Species abbreviation for the human or mouse databases ("HS" or "MM").
collection	Collection abbreviation, such as "H" or "C1". Use <code>msigdb_collections()</code> for the available options.
subcollection	Sub-collection abbreviation, such as "CGP" or "BP". Use <code>msigdb_collections()</code> for the available options.
category	[Deprecated] use the collection argument
subcategory	[Deprecated] use the subcollection argument

Details

Historically, the MSigDB resource has been tailored to the analysis of human-specific datasets, with gene sets exclusively aligned to the human genome. Starting with release 2022.1, MSigDB incorporated a database of mouse-native gene sets and was split into human and mouse divisions ("Hs" and "Mm"). Each one is provided in the approved gene symbols of its respective species. The versioning convention of MSigDB is in the format Year.Release.Species. The genes within each gene set may originate from a species different from the database target species as indicated by the `gs_source_species` and `db_target_species` fields.

Mouse MSigDB includes gene sets curated from mouse-centric datasets and specified in native mouse gene identifiers, eliminating the need for ortholog mapping.

To access the full dataset, please install the `msigdbdf` package (not available on CRAN):

```
install.packages("msigdbdf", repos = "https://igordot.r-universe.dev")
```

Value

A data frame of gene sets with one gene per row.

References

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

msigdbr_check_data *Check that the data package is installed*

Description

Check that the 'msigdbdf' data package is installed. If not, provide instructions for installation. A dependency listed in DESCRIPTION Suggests is not guaranteed to be installed.

Usage

```
msigdbr_check_data(require_data = TRUE)
```

Arguments

`require_data` Stop execution if the data package is not installed.

msigdb_collections *List the collections available in the msigdb package*

Description

List the collections available in the msigdb package

Usage

```
msigdb_collections(db_species = "Hs")
```

Arguments

db_species Species abbreviation for the human or mouse databases ("Hs" or "Mm").

Value

A data frame of the available collections.

msigdb_species *List the species available in the msigdb package*

Description

List the species available in the msigdb package

Usage

```
msigdb_species()
```

Value

A data frame of the available species.

Index

msigdb, [2](#)
msigdb_check_data, [3](#)
msigdb_collections, [4](#)
msigdb_species, [4](#)