Package ‘msigdbr’

July 3, 2025

Type Package

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

Version 25.1.0

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>, Liberzon et al. 2015 <doi:10.1016/j.cels.2015.12.004>, Castanza et al. 2023 <doi: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), curl, dplyr (>= 1.1.1), lifecycle, methods, rlang, tibble, tidyselect (>= 1.2.0), tools

Suggests knitr, rmarkdown, roxygen2, testthat

Config/Needs/website rmarkdown

Encoding UTF-8

RoxygenNote 7.3.2

NeedsCompilation no

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

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

Repository CRAN

Date/Publication 2025-07-03 19:10:02 UTC

Contents

`msigdbr` *Retrieve the gene sets data frame*

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

```
msigdbr(
  db_species = "HS",
  species = "human",
  collection = NULL,
  subcollection = NULL,
  category = deprecated(),
  subcategory = deprecated()
)
```

Arguments

`db_species`	Species abbreviation for the human or mouse databases (`"HS"` or `"MM"`).
`species`	Species name for output genes, such as `"Homo sapiens"` or `"Mus musculus"`. Both scientific and common names are acceptable. Use `msigdbr_species()` for the available options.
`collection`	Collection abbreviation, such as `"H"` or `"C1"`. Use `msigdbr_collections()` for the available options.
`subcollection`	Sub-collection abbreviation, such as `"CGP"` or `"BP"`. Use `msigdbr_collections()` 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.

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

Value

A tibble (a data frame with class `tibble::tbl_df`) of gene sets with one gene per row.

References

`https://www.gsea-msigdb.org/gsea/msigdb/index.jsp`

`msigdbr_collections` *List the collections available in the msigdbr package*

Description

List the collections available in the msigdbr package

Usage

```
msigdbr_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.

`msigdbr_species` *List the species available in the msigdbr package*

Description

List the species available in the msigdbr package

Usage

```
msigdbr_species()
```

Value

A data frame of the available species.

Index