# Package ‘EGSEAdata’

April 7, 2026

**Title** Gene set collections for the EGSEA package

**Version** 1.38.0

**Date** 15-03-2017

**Author** Monther Alhamdoosh, Yifang Hu and Gordon K. Smyth

**Maintainer** Monther Alhamdoosh `<m.hamdoosh@gmail.com>`

**Description** This package includes gene set collections that are used for the Ensemble of Gene Set Enrichment Analyses (EGSEA) method for gene set testing. It includes Human and Mouse versions of the MSidDB (Subramanian, et al. (2005) PNAS, 102(43):15545-15550) and GeneSetDB (Araki, et al. (2012) FEBS Open Bio, 2:76-82) collections.

**biocViews** ExperimentData, Homo_sapiens_Data, Mus_musculus_Data, Rattus_norvegicus_Data

**Depends** R (>= 3.4)

**License** file LICENSE

**LazyLoad** yes

**NeedsCompilation** no

**Suggests** EGSEA

**RoxygenNote** 5.0.1

**git_url** https://git.bioconductor.org/packages/EGSEAdata

**git_branch** RELEASE_3_22

**git_last_commit** b03a3b9

**git_last_commit_date** 2025-10-29

**Repository** Bioconductor 3.22

**Date/Publication** 2026-04-07

## Contents

---

## `EGSEAdata-package` *Gene Set Collections for the EGSEA package*

### Description

This package includes gene set collections that are used for the Ensemble of Gene Set Enrichment Analyses (EGSEA) method for gene set testing. It includes Human and Mouse versions of the MSidDB (Subramanian, et al. (2005) PNAS, 102(43):15545-15550) and GeneSetDB (Araki, et al. (2012) FEBS Open Bio, 2:76-82) collections.

### Details

While the gene set collections in MSigDB and GeneSetDB have different names and purposes, some of these collections overlap. For example, both databases contain a Gene Ontology collection but MSigDB's collection aimed for a higher level of abstraction for the GO terms.

### Author(s)

Monther Alhamdoosh, Yifang Hu and Gordon K. Smyth

### References

Monther Alhamdoosh, Milica Ng, Nicholas J. Wilson, Julie M. Sheridan, Huy Huynh, Michael J. Wilson, Matthew E. Ritchie; Combining multiple tools outperforms individual methods in gene set enrichment analyses. Bioinformatics 2017; 33 (3): 414-424. doi: 10.1093/bioinformatics/btw623

---

## `egsea.data` *EGSEAdata databases information*

### Description

It displays information about the available gene set collections in EGSEAdata for a species.

### Usage

```
egsea.data(species = "human", simple = FALSE, returnInfo = FALSE)
```

### Arguments

| | |
|---|---|
| `species` | character, a species name and used to retreive the number of gene sets for that particular species. Default is "human". Accepted values are "human", "homo sapiens", "hs", "mouse", "mus musculus", "mm", "rat", "rattus norvegicus" or "rn". |
| `simple` | logical, whether to display the number of gene sets in each collection or not. |
| `returnInfo` | logical, whether to print out the databases information or return it as a list. |

### Details

It prints out for each database: the database name, version, update/download date, data source, supported species, gene set collections, the names of the related R data objects and the number of gene sets in each collection.

### Value

nothing.

### Examples

```
# Example of egsea.data
egsea.data()
```

---

## `gsetdb.human` *GeneSetDB Human Collections*

### Description

Human gene set collections from the GeneSetDB

### Format

list

### Details

Procedure
1. The Human GMT file was downloaded from the website.
2. The gene set sources and categories were manually compiled from the Help page.
3. An R list was created for the gene set categories.
4. An annotation data frame was created for the gene sets.
5. An R data object was written using save().

### Source

Araki Hiromitsu,Knapp Christoph,Tsai Peter and Print Cristin(2012), GeneSetDB: A comprehensive meta-database, statistical and visualisation framework for gene set analysis, FEBS Open Bio, 2, doi: 10.1016/j.fob.2012.04.003 Downloaded from http://www.genesetdb.auckland.ac.nz/

### See Also

Invoke egsea.data() to find out the current version and latest download/update date.

---

## `gsetdb.mouse` *GeneSetDB Mouse Collections*

---

### Description

Mouse gene set collections from the GeneSetDB

### Format

list

### Details

Procedure
1. The Mouse GMT file was downloaded from the website.
2. The gene set sources and categories were manually compiled from the Help page.
3. An R list was created for the gene set categories.
4. An annotation data frame was created for the gene sets.
5. An R data object was written using save().

### Source

Araki Hiromitsu,Knapp Christoph,Tsai Peter and Print Cristin(2012), GeneSetDB: A comprehensive meta-database, statistical and visualisation framework for gene set analysis, FEBS Open Bio, 2, doi: 10.1016/j.fob.2012.04.003 Downloaded from http://www.genesetdb.auckland.ac.nz/

### See Also

Invoke egsea.data() to find out the current version and latest download/update date.

---

## `gsetdb.rat` *GeneSetDB Rat Collections*

---

### Description

Rat gene set collections from the GeneSetDB

### Format

list

### Details

Procedure
1. The Rat GMT file was downloaded from the website.
2. The gene set sources and categories were manually compiled from the Help page.
3. An R list was created for the gene set categories.
4. An annotation data frame was created for the gene sets.
5. An R data object was written using save().

## Source

Araki Hiromitsu,Knapp Christoph,Tsai Peter and Print Cristin(2012), GeneSetDB: A comprehensive meta-database, statistical and visualisation framework for gene set analysis, FEBS Open Bio, 2, doi: 10.1016/j.fob.2012.04.003 Downloaded from http://www.genesetdb.auckland.ac.nz/

## See Also

Invoke egsea.data() to find out the current version and latest download/update date.

---

# `il13.data` *Human IL-13 dataset*

## Description

The voom object calculated from the TMM normalized count matrix of RNA-seq performed on samples of human normal PBMCs, IL-13 stimulated PBMCs and IL-13R antagnonist PBBMCs. It also contains the contrast matrix of this experiment.

## Format

A List object with two components: voom and contra.

## Source

The count matrix of this experiment is vailable from the GEO database `www.ncbi.nlm.nih.gov/geo/` as series GSE79027.

---

# `il13.data.cnt` *Human IL-13 dataset - Raw Counts*

## Description

It contains the raw count matrix of RNA-seq performed on samples of human normal PBMCs, IL-13 stimulated PBMCs and IL-13R antagnonist PBBMCs. It also contains the contrast and design matrices of this experiment. The gene symbols mapping is also included.

## Format

A List object with five components: counts, group, design, contra and genes.

## Source

The FASTQ files of this experiment are vailable from the GEO database `www.ncbi.nlm.nih.gov/geo/` as series GSE79027.

---

## `il13.gsa` *EGSEA analysis results on the human IL-13 dataset*

---

### Description

EGSEA analysis was performed on the il13.data from the EGSEAdata package using the KEGG pathways, c2 and c5 gene set collections. Type show(il13.gsa) to see the version of datasets/packages that were used.

### Format

An object of class EGSEAResults

### Source

The dataset of this analysis is available in the EGSEAdata package.

---

## `kegg.pathways` *KEGG Pathways Collections*

---

### Description

Human, Mouse and Rat gene set collections from the KEGG database

### Format

list

### Details

The collections were generated using the following R script
library(gage)
kegg.pathways = list()
species.all = c("human", "mouse", "rat")
for (species in species.all)
kegg = kegg.gsets(species = species, id.type = "kegg")
kegg.pathways[[species]] = kegg

save(kegg.pathways, file='kegg.pathways.rda')

### Source

Luo, W., Friedman, M., Shedden K., Hankenson, K. and Woolf, P GAGE: Generally Applicable Gene Set Enrichment for Pathways Analysis. BMC Bioinformatics 2009, 10:161 Obtained from **gage** using the function `kegg.gsets()`

### See Also

Invoke egsea.data() to find out the current version and latest download/update date.

---

`mam.data` *Mouse mammary cell dataset*

---

## Description

The voom object calculated from TMM normalized count matrix of RNA-seq performed on samples of the epithelial cells of the mouse mammary glands from three populations: basal, luminal progenitor and mature luminal. It also contains the contrast matrix of this experiment.

## Format

A List object with two components: voom and contra.

## Source

The count matrix of this experiment is vailable from the GEO database `www.ncbi.nlm.nih.gov/geo/` as series GSE63310.

---

`Mm.c2` *Mouse C2 MSigDB Gene Set Collections*

---

## Description

Mouse orthologs gene set collections from the MSigDB database

## Format

list

## Details

Procedure
1. The current msigdb_vx.xml file was downloaded.
2. Human Entrez Gene IDs were mapped to Mouse Entrez Gene IDs, using the HGNC Comparison of Orthology Predictions (HCOP).
3. The collection was converted to a list in R, and written to a RData file using save().

## Source

Downloaded from http://bioinf.wehi.edu.au/software/MSigDB/. Extracted from Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles PNAS 2005 102 (43) 15545-15550

## See Also

Invoke egsea.data() to find out the current version and latest download/update date.

---

`Mm.c3` *Mouse C3 MSigDB Gene Set Collections*

---

### Description

Mouse orthologs gene set collections from the MSigDB database

### Format

list

### Details

Procedure
1. The current msigdb_vx.xml file was downloaded.
2. Human Entrez Gene IDs were mapped to Mouse Entrez Gene IDs, using the HGNC Comparison of Orthology Predictions (HCOP).
3. The collection was converted to a list in R, and written to a RData file using save().

### Source

Downloaded from http://bioinf.wehi.edu.au/software/MSigDB/. Extracted from Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles PNAS 2005 102 (43) 15545-15550

### See Also

Invoke egsea.data() to find out the current version and latest download/update date.

---

`Mm.c4` *Mouse C4 MSigDB Gene Set Collections*

---

### Description

Mouse orthologs gene set collections from the MSigDB database

### Format

list

### Details

Procedure
1. The current msigdb_vx.xml file was downloaded.
2. Human Entrez Gene IDs were mapped to Mouse Entrez Gene IDs, using the HGNC Comparison of Orthology Predictions (HCOP).
3. The collection was converted to a list in R, and written to a RData file using save().

## Source

Downloaded from http://bioinf.wehi.edu.au/software/MSigDB/ Extracted from Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles PNAS 2005 102 (43) 15545-15550

## See Also

Invoke egsea.data() to find out the current version and latest download/update date.

---

# `Mm.c5` *Mouse C5 MSigDB Gene Set Collections*

## Description

Mouse orthologs gene set collections from the MSigDB database

## Format

list

## Details

Procedure
1. The current msigdb_vx.xml file was downloaded.
2. Human Entrez Gene IDs were mapped to Mouse Entrez Gene IDs, using the HGNC Comparison of Orthology Predictions (HCOP).
3. The collection was converted to a list in R, and written to a RData file using save().

## Source

Downloaded from http://bioinf.wehi.edu.au/software/MSigDB/. Extracted from Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles PNAS 2005 102 (43) 15545-15550

## See Also

Invoke egsea.data() to find out the current version and latest download/update date.

## `Mm.c6` *Mouse C6 MSigDB Gene Set Collections*

### Description

Mouse orthologs gene set collections from the MSigDB database

### Format

list

### Details

Procedure
1. The current msigdb_vx.xml file was downloaded.
2. Human Entrez Gene IDs were mapped to Mouse Entrez Gene IDs, using the HGNC Comparison of Orthology Predictions (HCOP).
3. The collection was converted to a list in R, and written to a RData file using save().

### Source

Downloaded from http://bioinf.wehi.edu.au/software/MSigDB/. Extracted from Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles PNAS 2005 102 (43) 15545-15550

### See Also

Invoke egsea.data() to find out the current version and latest download/update date.

## `Mm.c7` *Mouse C7 MSigDB Gene Set Collections*

### Description

Mouse orthologs gene set collections from the MSigDB database

### Format

list

### Details

Procedure
1. The current msigdb_vx.xml file was downloaded.
2. Human Entrez Gene IDs were mapped to Mouse Entrez Gene IDs, using the HGNC Comparison of Orthology Predictions (HCOP).
3. The collection was converted to a list in R, and written to a RData file using save().

## Source

Downloaded from http://bioinf.wehi.edu.au/software/MSigDB/. Extracted from Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles PNAS 2005 102 (43) 15545-15550

## See Also

Invoke egsea.data() to find out the current version and latest download/update date.

---

# `Mm.H` *Mouse H MSigDB Gene Set Collections*

## Description

Mouse orthologs gene set collections from the MSigDB database

## Format

list

## Details

Procedure
1. The current msigdb_vx.xml file was downloaded.
2. Human Entrez Gene IDs were mapped to Mouse Entrez Gene IDs, using the HGNC Comparison of Orthology Predictions (HCOP).
3. The collection was converted to a list in R, and written to a RData file using save().

## Source

Downloaded from http://bioinf.wehi.edu.au/software/MSigDB/. Extracted from Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles PNAS 2005 102 (43) 15545-15550

## See Also

Invoke egsea.data() to find out the current version and latest download/update date.

## `msigdb` *MSigDB Gene Set Collections*

### Description

Gene set collections from the MSigDB database

### Format

list

### Details

Procedure
1. The current msigdb_vx.xml file was downloaded.
2. It was parsed using xmlParse() and then converted to list using xmlToList() 3. The resulting list was written to an RData file using save().
This dataset is mainly used to extract MSigDB gene set annotation and the human gene set collections.

### Source

Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles PNAS 2005 102 (43) 15545-15550

### See Also

Invoke egsea.data() to find out the current version and latest download/update date.

# Index