

Package ‘gskb’

October 12, 2020

Type Package

Title Gene Set data for pathway analysis in mouse

Version 1.20.0

Date 2015-03-18

Author Valerie Bares, Xijin Ge

Maintainer Valerie Bares <valerie.bares@sdstate.edu>

Description Gene Set Knowledgebase (GSKB) is a comprehensive knowledgebase for pathway analysis in mouse. Interpretation of high-throughput genomics data based on biological pathways constitutes a constant challenge, partly because of the lack of supporting pathway database. We created a functional genomics knowledgebase in mouse, which includes 33,261 pathways and gene sets compiled from 40 sources such as Gene Ontology, KEGG, GeneSetDB, PANTHER, microRNA and transcription factor target genes, etc. In addition, we also manually collected and curated 8,747 lists of differentially expressed genes from 2,526 published gene expression studies to enable the detection of similarity to previously reported gene expression signatures. These two types of data constitute a comprehensive Gene Set Knowledgebase (GSKB), which can be readily used by various pathway analysis software such as gene set enrichment analysis (GSEA).

As a first step, we gathered annotation information from 40 existing databases for mouse-related gene sets. These gene sets are divided into 7 categories, namely, Gene Ontology, Curated pathways, Metabolic Pathways, Transcription Factor (TF) and microRNA target genes, location (cytogenetics band), and others. We used information in GeneSetDB for some of the databases. Detailed information on these 40 sources and the citations is available

<http://ge-lab.org/gskb/Table%201-sources.pdf>.

The gene lists from literature were retrieved manually from individual gene expression studies through a process similar to the one used to create AraPath, a similar resource for Arabidopsis[12]. As most expression studies upload raw data to repositories like GEO and ArrayExpress, we used the meta-data in these databases to search for publications. We scanned all datasets we can found and retrieved 4,313 potentially useful papers reporting gene expression studies in mouse. These papers were individually read by curators to identify lists of differentially expressed genes in various conditions. We compiled a total of 8,747 lists of differently expressed genes from 2,518 of

papers. Each gene list was annotated with a unique name, brief description, and publication information, similar to the protocol used in MSigDB and Arapath. These gene lists constitute a large collection of published gene expression signatures that form a foundation for interpret new gene lists and expression profiles.

More information about this data is available here

<http://ge-lab.org/gskb/>. There is also a paper describing these data are currently in revision by Database: The Journal of Biological Databases and Curation.

License Artistic-2.0

biocViews ExperimentData, Mus_musculus

Depends R (>= 3.2.0)

Suggests PGSEA

git_url <https://git.bioconductor.org/packages/gskb>

git_branch RELEASE_3_11

git_last_commit ad0d901

git_last_commit_date 2020-04-27

Date/Publication 2020-10-12

R topics documented:

mm_GO	2
mm_location	3
mm_metabolic	3
mm_miRNA	4
mm_other	4
mm_pathway	5
mm_TF	5
Index	6

mm_GO

Gene Ontology Data for Mouse

Description

Gene Set Knowledgebase (GSKB) is a comprehensive knowledgebase for pathway analysis in mouse. This object contains genes sets derived from Gene Ontology database. This data can be used by programs such as GSEA and pGSEA.

Usage

```
data(mm_GO)
```

Value

A list is returned, where each list element is a gene set.

References

Liming Lai, Jason Hennessey, Yuguang Ban, Wei Wang, Jianli Qi, Gaixin Jiang, Valerie Bares, and Steven Xijin Ge, GSKB: A comprehensive knowledgebase for pathway analysis in mouse. Database: The journal of biological databases and curation (in revision).

mm_location

Chromosomal Location Data for Mouse

Description

Gene Set Knowledgebase (GSKB) is a comprehensive knowledgebase for pathway analysis in mouse. This object contains lists of genes grouped by location on chromosomes. This data can be used by programs such as GSEA and pGSEA.

Usage

```
data(mm_location)
```

Value

A list is returned, where each list element is a gene set.

References

Liming Lai, Jason Hennessey, Yuguang Ban, Wei Wang, Jianli Qi, Gaixin Jiang, Valerie Bares, and Steven Xijin Ge, GSKB: A comprehensive knowledgebase for pathway analysis in mouse. Database: The journal of biological databases and curation (in revision).

mm_metabolic

Metabolic Pathways Data for Mouse

Description

Gene Set Knowledgebase (GSKB) is a comprehensive knowledgebase for pathway analysis in mouse. This object contains metabolic pathways based on KEGG, EHMN, MouseCyc and REACTOME. This data can be used by programs such as GSEA and pGSEA.

Usage

```
data(mm_metabolic)
```

Value

A list is returned, where each list element is a gene set.

References

Liming Lai, Jason Hennessey, Yuguang Ban, Wei Wang, Jianli Qi, Gaixin Jiang, Valerie Bares, and Steven Xijin Ge, GSKB: A comprehensive knowledgebase for pathway analysis in mouse. Database: The journal of biological databases and curation (in revision).

mm_miRNA

miRNA Target Genes Data for Mouse

Description

Gene Set Knowledgebase (GSKB) is a comprehensive knowledgebase for pathway analysis in mouse. This object contains information about target genes of microRNAs based on information from miRDB, TarBase, miRTarBase etc. This data can be used by programs such as GSEA and pGSEA.

Usage

```
data(mm_miRNA)
```

Value

A list is returned, where each list element is a gene set.

References

Liming Lai, Jason Hennessey, Yuguang Ban, Wei Wang, Jianli Qi, Gaixin Jiang, Valerie Bares, and Steven Xijin Ge, GSKB: A comprehensive knowledgebase for pathway analysis in mouse. Database: The journal of biological databases and curation (in revision).

mm_other

Other Data for Mouse

Description

Gene Set Knowledgebase (GSKB) is a comprehensive knowledgebase for pathway analysis in mouse. This object contains gene sets based on various sources such as STITCH (interaction networks of chemicals and proteins), MPO (mammalian phenotype ontology) etc. This data can be used by programs such as GSEA and pGSEA.

Usage

```
data(mm_other)
```

Value

A list is returned, where each list element is a gene set.

References

Liming Lai, Jason Hennessey, Yuguang Ban, Wei Wang, Jianli Qi, Gaixin Jiang, Valerie Bares, and Steven Xijin Ge, GSKB: A comprehensive knowledgebase for pathway analysis in mouse. Database: The journal of biological databases and curation (in revision).

`mm_pathway`*Pathway Data for Mouse*

Description

Gene Set Knowledgebase (GSKB) is a comprehensive knowledgebase for pathway analysis in mouse. This object contains curated pathways based on Biocarta, PANTHER, WikiPathways, IONH, and NetPath. This data can be used by programs such as GSEA and pGSEA.

Usage

```
data(mm_pathway)
```

Value

A list is returned, where each list element is a gene set.

References

Liming Lai, Jason Hennessey, Yuguang Ban, Wei Wang, Jianli Qi, Gaixin Jiang, Valerie Bares, and Steven Xijin Ge, GSKB: A comprehensive knowledgebase for pathway analysis in mouse. Database: The journal of biological databases and curation (in revision).

`mm_TF`*Transcription Factor Target Genes Data for Mouse*

Description

Gene Set Knowledgebase (GSKB) is a comprehensive knowledgebase for pathway analysis in mouse. This object contains gene sets corresponding to target genes of transcription factors. Some gene sets are based on experimentally confirmed data, while others are based on predictions. These are based on TFacts, TRED, CircuitsDB, and TRANSFAC.

Usage

```
data(mm_TF)
```

Value

A list is returned, where each list element is a gene set.

References

Liming Lai, Jason Hennessey, Yuguang Ban, Wei Wang, Jianli Qi, Gaixin Jiang, Valerie Bares, and Steven Xijin Ge, GSKB: A comprehensive knowledgebase for pathway analysis in mouse. Database: The journal of biological databases and curation (in revision).

Index

* datasets

- mm_GO, 2
- mm_location, 3
- mm_metabolic, 3
- mm_miRNA, 4
- mm_other, 4
- mm_pathway, 5
- mm_TF, 5

- mm_GO, 2
- mm_location, 3
- mm_metabolic, 3
- mm_miRNA, 4
- mm_other, 4
- mm_pathway, 5
- mm_TF, 5