Package 'yeastGSData'

February 26, 2026

Title Yeast Gold Standard Data

Version 0.48.0

Author R. Gentleman

Description A collection of so-called gold (and other) standard data sets

biocViews ExperimentData, Saccharomyces_cerevisiae_Data

Maintainer R. Gentleman <rgentlem@fhcrc.org>

License Artistic-2.0

git_url https://git.bioconductor.org/packages/yeastGSData

git_branch RELEASE_3_22

git_last_commit e3700d6

git_last_commit_date 2025-10-29

Repository Bioconductor 3.22

Date/Publication 2026-02-26

Contents

BayesFactorNGS   *The Binary Negative Gold Standard Developed via Bayes Factor*

Description

This data consists of the 7739 binary interactions with computed Bayes factor of less than -3. This dataset was created 16 June 2009.

Usage

```
data(BayesFactorNGS)
```

Format

A data frame with observations on the following 5 variables.

Orf1 The ORF for gene 1

Orf2 The ORF for gene 2

Tested The number of times the interaction was tested

Observed The number of times the interaction was found

log_BF The log of the resulting Bayes factor

Details

None.

Examples

```
data(BayesFactorNGS)
```

BayesFactorPGS   *The Binary Positive Gold Standard Developed via Bayes Factor*

Description

This data consists of the 10200 binary interactions with computed Bayes factor of greater than 3. This dataset was created 16 June 2009.

Usage

```
data(BayesFactorPGS)
```

Format

A data frame with observations on the following 5 variables.

Orf1 The ORF for gene 1

Orf2 The ORF for gene 2

Tested The number of times the interaction was tested

Observed The number of times the interaction was found

log_BF The log of the resulting Bayes factor

Details

None.

Examples

```
data(BayesFactorPGS)
```

BinaryGS   *The Binary Gold Standard Data set Reported by Yu et al*

Description

This data consists of the 1318 binary interactions Yu et al reported as their binary gold standard data set.

Usage

```
data(BinaryGS)
```

Format

A data frame with 1318 observations on the following 2 variables.

ORF1 The ORF for gene 1

ORF2 The ORF for gene 2

Details

None.

Source

The data were downloaded from <http://interactome.dfci.harvard.edu/S_cerevisiae> on Nov 21, 2008.

References

H. Yu et al, High Quality Binary Protein Interaction Map of the Yeast Interactome Network, Science 322, 104, 2008.

Examples

```
data(BinaryGS)
```

MIPSGS	*MIPS Gold Standard Protein Interactions*

Description

The MIPS gold standard protein complex data set downloaded from the Gerstein Lab web site.

Usage

```
data(MIPSGS)
```

Format

A data frame with 8617 observations on the following 5 variables.

ORF1 The ORF for one gene

ORF2 The ORF for the second gene

CID The MIPS protein complex ID

NAME The name of the complex.

NUMBER The number of proteins in the complex.

Details

The data are essentially multiprotein complexes, curated from MIPS data, see also the data set in Mpact, which is related.

The data are all pairwise members of each complex.

Yu et al. state " To compile a reference data set with the lowest false-positive rate, we consider two proteins as interaction part- ners if and only if they are in the same complex of the highest level in the catalog. "

Source

<http://interolog.gersteinlab.org>

References

Annotation Transfer Between Genomes: Protein-Protein Interologs and Protein-DNA Regulogs, H. Yu et al, Genome Research, 1107-1118, 2004.

Examples

```
data(MIPSGS)
```

Mpact   *Protein Interaction Data from Mpact*

Description

The data are protein interactions, downloaded from MPACT. They are stored in a large character matrix, with seven columns, describing the interactions.

Usage

```
data(Mpact)
```

Format

The data are stored as a matrix, with columns

- ORF1The ORF for gene 1
- GENE1The symbol for gene 1
- ORF2The ORF for gene 2
- GENE2The symbol for gene 2
- DESCRA description of one, or both genes
- EVIAn evidence code.

Details

It is unlikely that the variables GENE1 and GENE2 can be relied on, as names change, so ORF1 and ORF2 should be preferred, and even these should be compared to current databases to see if they have been supplanted.

The DESCR field is incomplete, and seems to be inconsistent. It would probably better to rely on the the ORFs to obtain documentation on the ORFs from other sources.

The EVI variable, gives one, or more evidence codes, separate by commas ,. The evidence codes are further detailed in the MpactEvidenceCodes data object. Evidence codes can be helpful in filtering out interactions that might give rise to circularity in an analysis. By that we mean, that if you are analyzing data that comes from one of the experiments that was used to establish this gold standard data set, it might be best to filter those interactions out. You should be careful to only filter them, if their only evidence is from the experiment you are analyzing (if there is other evidence for the interaction it should be retained).

Source

The data were downloaded from <ftp://ftpmips.gsf.de/yeast/PPI/>.

References

Guldener U, Munsterkotter M, Oesterheld M, Pagel P, Ruepp A, Mewes HW, Stumpflen V(2006). MPact: the MIPS protein interaction resource on yeast. Nucl. Acids Res. 2006 34: D436-D441 PMID: 16381906

See Also

MpactEvidenceCodes

Examples

```
data(Mpact)
Mpact[1:3,]
```

MpactEvidenceCodes   *MIPS Evidence Codes*

Description

The data in `Mpact` are interaction data from MIPS. Each interaction has one or more evidence code, that is intended to document the basis on which an interaction is presumed.

Usage

```
data(MpactEvidenceCodes)
```

Format

A character vector of the descriptions, with names given by the evidence codes.

Details

There is a nesting in the evidence codes that is not directly reflected in this data. The first three names are 901, 901.01 and 901.01.01, so the first is a top level term, the second is nested under it, and the third under the second.

Source

The data were downloaded from `ftp://ftpmips.gsf.de/yeast/PPI/`.

References

Guldener U, Munsterkotter M, Oesterheld M, Pagel P, Ruepp A, Mewes HW, Stumpflen V(2006). MPact: the MIPS protein interaction resource on yeast. Nucl. Acids Res. 2006 34: D436-D441 PMID: 16381906

See Also

`Mpact`

Examples

```
data(MpactEvidenceCodes)
MpactEvidenceCodes[1:3]
```

NEGGS *A Gold Standard for Negative Interactions*

Description

These data were supplied as supplementary material, for the paper below, as a data set for negative interactions.

Usage

```
data(NEGGS)
```

Format

A data frame with 2708746 observations on the following 4 variables.

ORF1 The ORF of one interactor.

ORF2 The ORF of the second interactor.

LOC1 A description of where the first interactor is (typically) located in the cell.

LOC2 A description of where the first interactor is (typically) located in the cell.

Details

The data are potentially problematic, since not being in the same cellular component does not mean that two proteins will not interact in some particular assay.

Only a very broad grouping of location is given, and one may want to refer to a more recent and potentially more authorative source.

Source

<http://interolog.gersteinlab.org>

References

Annotation Transfer Between Genomes: Protein-Protein Interologs and Protein-DNA Regulogs, H. Yu et al, Genome Research, 1107-1118, 2004.

Examples

```
data(NEGGS)
table(NEGGS$LOC1)
table(NEGGS$LOC2)
```

POSPS   *A Platinum Standard Data Set*

Description

While Yu et al. call this a platinum standard data set, it is really a gold standard data set for binary physical interactions.

Usage

```
data(POSPS)
```

Format

A data frame with 1867 observations on the following 2 variables.

ORF1 a character vector

ORF2 a character vector

Details

These data, reported in the paper below, are intended to represent well established binary physical interactions between proteins. This contrasts with the *gold standard* MIPSGS which describes multi-protein complexes.

Ye et al, describe the construction as follows: " Briefly, the data set contains physical interactions from complex protein structures in the Protein Data Bank (Westbrook et al. 2003), verified interactions from small-scale experiments (Mewes et al. 2000; Xenarios et al. 2002; Bader et al. 2003), and protein pairs from small MIPS catalog complexes (4 or fewer subunits)."

References

Annotation Transfer Between Genomes: Protein-Protein Interologs and Protein-DNA Regulogs, H. Yu et al, Genome Research, 1107-1118, 2004.

Examples

```
data(POSPS)
```

PRS   *The Positive Reference Set Reported by Yu et al*

Description

These data consist of the 188 binary interactions that Yu et al curated and referred to as the positive reference set.

Usage

```
data(PRS)
```

Format

A data frame with 188 observations on the following 2 variables.

ORF1 The ORF for gene 1.

ORF2 The ORF for gene 2.

Details

None.

Source

The data were downloaded from http://interactome.dfci.harvard.edu/S_cerevisiae on Nov 21, 2008.

References

H. Yu et al, High Quality Binary Protein Interaction Map of the Yeast Interactome Network, Science 322, 104, 2008.

Examples

```
data(PRS)
## maybe str(PRS) ; plot(PRS) ...
```

Index