### Description

List of complex estimates after filtering baits prone to systematic bias

### Usage

```r
data(gavin06FilteredEstimates)
data(krogan06FilteredEstimates)
```

### Details

`gavin06FilteredEstimates` contains results from an analysis of the AP-MS data published by Gavin et al. (2006). These estimates were constructed using `findComplexes` with a sensitivity parameter of .70, specificity of .999, and `commonFrac=0.5`.

`krogan06FilteredEstimates` contains results from an analysis of the AP-MS data published by Krogan et al. (2006). These estimates were constructed using `findComplexes` with a sensitivity parameter of .70, specificity of .999, and `commonFrac=1/3`.

Both sets of estimates are reported as lists of length three, corresponding to MBME, SBMH, and UnRBB complex estimate types (see Scholtens et al., 2005). Each of the three elements contains a list of character vectors of estimated complex members.

### Source


### References

Gavin, et al. (2006)

Krogan, et al. (2006)
See Also

gavinBP2006, kroganBPMat2006, \code{findComplexes}

Examples

data(gavin06FilteredEstimates)
lapply(gavin06FilteredEstimates, FUN = length)

data(krogan06FilteredEstimates)
lapply(krogan06FilteredEstimates, FUN = length)

---

**HMSPCI**

*High-Throughput Mass Spectromic Protein Complex Identification (HMS-PCI) Data from Ho, et al. (2002)*

**Description**

HMS-PCI data published by Ho, et al. (2002).

**Usage**

data(HMSPCI)
data(HMSPCIgraph)

**Details**

HMSPCI is a matrix of the HMS-PCI data published by Ho, et al. (2002). The 493 rows correspond to bait proteins and the 1578 columns correspond to proteins found as hits in the HMSPCI experiment and are named accordingly. The first 493 column names are the same as the 493 row names; bait proteins can also be found as hits by other baits, hence their inclusion as columns in HMSPCI. An entry of "1" in the ith row and jth column of HMSPCI indicates that bait protein i found protein j as a hit. All other entries are "0". There are a total of 3687 "1" entries in the matrix, corresponding to the 3687 comemberships detected in the experiment.

HMSPCIgraph is a graphNEL object in which 1578 nodes represent proteins and 3687 directed edges represent comemberships detected in the purification. Each directed edge originates at the bait and ends at the hit. Each edge in HMSPCIgraph corresponds to an entry of "1" in the HMSPCI matrix.

These data are available at http://www.mdsp.com/yeast.

**Source**


**References**


See Also

MBMEchHSMPCL, SBMHchHSMPCL, UnRBBchHSMPCL, TAP

Examples

data(HMSPCI)
which(HMSPCI["YAL015C",]==1)

data(HMSPCIgraph)
adj(HMSPCIgraph,"YAL015C")

**Krogan**

*High-Definition Macromolecular Composition of Yeast RNA-Processing Complexes (2004)*

Description

AP-MS data published by Krogan et al. (2004)

Usage

data(Krogan)

Details

*Krogan* is a matrix of the AP-MS data published by Krogan et al (2004). The 153 rows correspond to bait proteins and the 485 columns correspond to proteins found as hits in the AP-MS experiment and are named accordingly. The first 153 column names are the same as the 153 row names; bait proteins can also be found as hits by other baits, hence their inclusion as columns in Krogan. An entry of "1" in the i\(^{th}\) row and j\(^{th}\) column of Krogan indicates that bait protein i found protein j as a hit. All other entries are "0". There are a total of 1132 "1" entries in the matrix, corresponding to the 1132 comemberships detected in the experiment.

These data are available at http://www.molecule.org/cgi/content/full/13/2/225/DC1/.

Source


References


See Also

MBMEcKrogan
LCjoin

Computation in LxC measure

**Examples**

data(Krogan)

**Description**

Computes the change in the P=LxC measure for AP-MS protein data when two protein complex estimates are combined into one complex.

**Details**

These functions are used to evaluate changes in the penalized likelihood when two complexes are combined. They are not meant to be directly used.

**Value**

The numeric value of the change in P=LxC when two complexes are combined.

**Author(s)**

Denise Scholtens

**References**


**See Also**

bhmaxSubgraph, mergeComplexes, findComplexes

**Examples**

data(apEX)
PCMG0 <- bhmaxSubgraph(apEX)
PCMG1 <- mergeComplexes(PCMG0, apEX, sensitivity=.7, specificity=.75)
Description

Affiliation matrices with rows corresponding to proteins and columns corresponding to complexes.

Usage

data(MBMEcHMSPCI)
data(SBMHcHMSPCI)
data(UnRBBcHMSPCI)

Details

These are the results from an analysis of the HMS-PCI data (Ho et al., 2002) described by Scholtens and Gentleman (2004) and Scholtens, Vidal, and Gentleman (submitted). These estimates were constructed using `findComplexes` with a sensitivity parameter of .75, specificity of .99, and **Beta**=-0.2 for externally derived similarity measure based on Gene Ontology cellular component annotation (see Scholtens and Gentleman (2004)).


Source


References


See Also

`HMSPCI`, `HMSPCIgraph`, `code(findComplexes)`

Examples

data(MBMEcHMSPCI)
MBMEcHMSPCI[1:4,1:4]
which(MBMEcHMSPCI[,"MBME1"]==1)
TAP complexes

Krogan complexes  

**Description**

Affiliation matrices with rows corresponding to proteins and columns corresponding to complexes.

**Usage**

```r
data(MBMEcKrogan)
```

**Details**

These are the results from an analysis of the AP-MS data (Krogan et al., 2004). These estimates were constructed using `findComplexes` with a sensitivity parameter of .75, specificity of .99, and Beta=-0.2 for externally derived similarity measure based on Gene Ontology cellular component annotation (see Scholtens and Gentleman (2004)).

`MBMEcHMSPC` contains 82 multi-bait-multi-edge complex estimates.

**Source**


**References**

High-Definition Macromolecular Composition of Yeast RNA-Processing Complexes; Molecular Cell, Vol 13, 225-239, 30 January 2004

**Examples**

```r
data(MBMEcKrogan)
MBMEcKrogan[1:4,1:4]
```

TAP complexes  

**TAP data complex estimates**

**Description**

Affiliation matrices with rows corresponding to proteins and columns corresponding to complexes.

**Usage**

```r
data(MBMEcTAP)
data(SBMHcTAP)
data(UnRBBcTAP)
```
Details

These are the results from an analysis of the TAP data (Gavin et al., 2002) by Scholtens and Gentleman (2004) and Scholtens, Vidal, and Gentleman (submitted). These estimates were constructed using findComplexes with a sensitivity parameter of .75, specificity of .995, and Beta=-0.2 for externally derived similarity measure based on Gene Ontology cellular component annotation (see Scholtens and Gentleman (2004)).


Source


References


See Also

TAP,TAPgraph,
\code{TAPgraph},
\code{findComplexes}

Examples

data(MBMEcTAP)
MBMEcTAP[1:10,1:3]
which(MBMEcTAP[,"MBME1"]==1)

TAP  Tandem Affinity Purification (TAP) Data from Gavin et al. (2002)

Description


Usage

data(TAP)
data(TAPgraph)
Details

TAP is a matrix of the TAP data published by Gavin, et al. (2002). The 455 rows correspond to bait proteins and the 1364 columns correspond to proteins found as hits in the TAP experiment and are named accordingly. The first 455 column names are the same as the 455 row names; bait proteins can also be found as hits by other baits, hence their inclusion as columns in TAP. An entry of "1" in the ith row and jth column of TAP indicates that bait protein i found protein j as a hit. All other entries are "0". There are a total of 3420 "1" entries in the matrix, corresponding to the 3420 comemberships detected in the experiment.

TAPgraph is a graphNEL object in which 1364 nodes represent proteins and 3420 directed edges represent comemberships detected in the purification. Each directed edge originates at the bait and ends at the hit. Each edge in TAPgraph corresponds to an entry of "1" in the TAP matrix.

These data are available in Supplementary Table S1 of Gavin et al. at http://www.nature.com.

Source


References


See Also

yTAP, MBMEcTAP, SBMHcTAP, JnRBBcTAP, HMSPCI

Examples

data(TAP)
which(TAP["Abd1",] ==1)

data(TAPgraph)
adj(TAPgraph,"Abd1")

---

apEX

Example data set for apComplex package

Description

A matrix and corresponding graph of AP-MS purifications from a small hypothetical experiment.

Usage

data(apEX)
data(apEXG)
Details

apEX is a matrix of hypothetical AP-MS data. The 4 rows correspond to bait proteins and the 8 columns correspond to proteins found as hits in the hypothetical experiment and are named accordingly. The first 4 column names are the same as the 4 row names; bait proteins can also be found as hits by other baits, hence their inclusion as columns in apEX. An entry of "1" in the ith row and jth column of apEX indicates that bait protein i found protein j as a hit. All other entries are "0". There are a total of 9 "1" entries in the matrix, corresponding to 9 comemberships detected in the experiment.

apEXG is a graphNEL object in which 8 nodes represent proteins and 9 directed edges represent comemberships detected in the purification. Each directed edge originates at the bait and ends at the hit. Each edge in apEXG corresponds to an entry of "1" in the apEX matrix.

These data are used in the apComplex vignette.

Examples

data(apEX)
apEX
data(apEXG)
adj(apEXG,"P1")

bhmaxSubgraph  
Find maximal BH-complete subgraph

Description

Given an adjacency matrix of bait-hit AP-MS protein data, this function finds the maximal BH-complete subgraphs and reports them as an affiliation matrix.

Usage

bhmaxSubgraph(adjMat, VBs=NULL, VPs=NULL, unrecip=1)

Arguments

adjMat  adjMat is an N by (N+M) adjacency matrix with N equal to the number of bait proteins and M equal to the number of hit-only proteins. adjMat should have row and column names corresponding to the proteins in the experiment. An entry of "1" in the ith row and jth column of adjMat corresponds to bait protein i finding protein j as a hit. All other entries should be 0.

VBs  VBs is an optional vector of viable baits.

VPs  VPs is an optional vector of viable prey.

unrecip  By default set to 1 so that unreciprocated bait-bait edges are treated as present. If set to 0, unreciprocated bait-bait edges will be treated as absent.
Details
A BH-complete subgraph with n bait nodes and m hit-only nodes for AP-MS data is defined as a subgraph for which all n*(n-1)+nm directed edges exist. A maximal BH-complete subgraph is a BH-complete subgraph which is not contained in any other BH-complete subgraph.

If VBs and/or VPs are not specified, then by default VBs will be assigned the set of baits that detect at least one prey and VPs the set of prey that are detected by at least one bait.

By default, unreciprocated bait-bait observations will be treated as present. If unrecip is set to 0, they will be treated as absent. If the sensitivity of the AP-MS technology is believed to be less than the specificity, then it is suggested that unrecip=1.

This function calls maxCliques from the RBGL package.

Value
A list of length one named ‘maxCliques’ which is itself a list of character vectors containing the names of the elements in the cliques.

Author(s)
Denise Scholtens

References


See Also
mergeComplexes, findComplexes

Examples

data(apEX)
PCMG0 <- bhmaxSubgraph(apEX)
PCMG1 <- mergeComplexes(PCMG0, apEX, sensitivity=.7, specificity=.75)

findComplexes Estimate a Protein Complex Membership Graph (PCMG) using protein complex comembership data from AP-MS technology

Description
Performs all steps in the local modeling algorithm described by Scholtens and Gentleman (2004) and Scholtens, Vidal, and Gentleman (submitted), beginning with an adjacency matrix recording bait-hit AP-MS data.
Usage

findComplexes(adjMat, VBs=NULL, VPs=NULL, simMat=NULL, sensitivity=.75, specificity=.995, Beta=0, commonFrac=2/3, wsVal = 2e7)

Arguments

adjMat Adjacency matrix of bait-hit data from an AP-MS experiment. Rows correspond to baits and columns to hits.
VBs VBs is an optional vector of viable baits.
VPs VPs is an optional vector of viable prey.
simMat An optional square matrix with entries between 0 and 1. Rows and columns correspond to the proteins in the experiment, and should be reported in the same order as the columns of adjMat. Higher values in this matrix are interpreted to mean higher similarity for protein pairs.
sensitivity Believed sensitivity of AP-MS technology.
specificity Believed specificity of AP-MS technology.
Beta Optional additional parameter for the weight to give data in simMat in the logistic regression model.
commonFrac This is the fraction of baits that need to be overlapping for a complex combination to be considered.
wsVal A numeric. This is the value assigned as the work-space in the call to fisher.test

Details

findComplexes performs all steps in the complex estimation algorithm using the apComplex package functions bhmaxSubgraph, LCdelta, and mergeComplexes. These steps can also be performed separately by the user.

If VBs and/or VPs are not specified, then by default VBs will be assigned the set of baits that detect at least one prey and VPs the set of prey that are detected by at least one bait.

By default commonFrac is set relatively high at 2/3. This means that some potentially reasonable complex combinations could be missed. For smaller data sets, users may consider decreasing the fraction. For larger data sets, this may cause a large increase in computation time.

Value

A list of character vectors containing the names of the proteins in the estimated complexes.

Author(s)

Denise Scholtens

References

See Also

bhmaxSubgraph, mergeComplexes

Examples

data(apEX)
PCMG2 <- findComplexes(apEX, sensitivity=.7, specificity=.75)

gavinBP2006

Tandem Affinity Purification (TAP) Data from Gavin et al. (2006)

Description


Usage

data(gavinBP2006)

Details

gavinBP2006 is a matrix of the TAP data published by Gavin, et al. (2006). The 1752 rows
 correspond to bait proteins and the 2551 columns correspond to proteins found as hits in the TAP
 experiment and are named accordingly. The first 1752 column names are the same as the 1752 row
 names; bait proteins can also be found as hits by other baits, hence their inclusion as columns in
 gavinBP2006. An entry of "1" in the ith row and jth column of gavinBP2006 indicates that
 bait protein i found protein j as a hit. All other entries are "0". There are a total of 19105 "1" entries
 in the matrix, corresponding to the 19105 comemberships detected in the experiment.

These data are available in the IntAct repository under Gavin et al - 2006.

Source

Gavin, et al. Functional Organization of the Yeast Proteome by Systematic Analysis of the Yeast

See Also

yTAP, MBMEcTAP, SBMHcTAP, UnRBBcTAP, HMSPCI

Examples

data(gavinBP2006)
which(gavinBP2006[1,]==1)
Tandem Affinity Purification (TAP) Data from Krogan et al. (2006)

Description

Usage
```r
data(kroganBPMat2006)
```

Details
`kroganBPMat2006` is a matrix of the TAP data published by Krogan, et al. (2006). The 2264 rows correspond to bait proteins and the 5361 columns correspond to proteins found as hits in the TAP experiment and are named accordingly. The first 2264 column names are the same as the 2264 row names; bait proteins can also be found as hits by other baits, hence their inclusion as columns in `kroganBPMat2006`. An entry of "1" in the ith row and jth column of `kroganBPMat2006` indicates that bait protein i found protein j as a hit. All other entries are "0". There are a total of "1" entries in the matrix, corresponding to the 63360 comemberships detected in the experiment.

These data were obtained from the Primary Source -

Source

See Also
`yTAP,MBMEcTAP,SBMHcTAP,UnRBBcTAP,HMSPCI`

Examples
```r
data(kroganBPMat2006)
which(kroganBPMat2006[1,]==1)
```

mergeComplexes

Iteratively combine columns in initial PCMG estimate

Description
Repeatedly applies the function `LCdelta` to make combinations of columns in the affiliation matrix representing the protein complex membership graph (PCMG) for AP-MS data.

Usage
```r
mergeComplexes(bhmax,adjMat,VBs=NULL,VPs=NULL,simMat=NULL,sensitivity=.75,specificity=.995,Beta=0,commonFrac=2/3,wsVal = 2e7)
```
mergeComplexes

Arguments

- **bhmax**: Initial complex estimates coming from bhmaxSubgraph
- **adjMat**: Adjacency matrix of bait-hit data from an AP-MS experiment. Rows correspond to baits and columns to hits.
- **VBs**: VBs is an optional vector of viable baits.
- **VPs**: VPs is an optional vector of viable prey.
- **simMat**: An optional square matrix with entries between 0 and 1. Rows and columns correspond to the proteins in the experiment, and should be reported in the same order as the columns of adjMat. Higher values in this matrix are interpreted to mean higher similarity for protein pairs.
- **sensitivity**: Beloved sensitivity of AP-MS technology.
- **specificity**: Believed specificity of AP-MS technology.
- **Beta**: Optional additional parameter for the weight to give data in simMat in the logistic regression model.
- **commonFrac**: This is the fraction of baits that need to be overlapping for a complex combination to be considered.
- **wsVal**: A numeric. This is the value assigned to the work-space in the call to fisher.test.

Details

The local modeling algorithm for AP-MS data described by Scholtens and Gentleman (2004) and Scholtens, Vidal, and Gentleman (2005) uses a two-component measure of protein complex estimate quality, namely $P = LxC$. Columns in cMat represent individual complex estimates. The algorithm works by starting with a maximal BH-complete subgraph estimate of cMat, and then improves the estimate by combining complexes such that $P = LxC$ increases.

By default commonFrac is set relatively high at 2/3. This means that some potentially reasonable complex combinations could be missed. For smaller data sets, users may consider decreasing the fraction. For larger data sets, this may cause a large increase in computation time.

Value

A list of character vectors containing the names of the proteins in the estimated complexes.

Author(s)

Denise Scholtens

References


See Also

bhmaxSubgraph, findComplexes
**plotComplex**

*Render complex estimates*

**Examples**

```r
data(apEX)
PCMG0 <- bhmaxSubgraph(apEX)
PCMG1 <- mergeComplexes(PCMG0, apEX, sensitivity=.7, specificity=.75)
```

**Description**

`plotComplex` renders complex estimates from the `apComplex` algorithm using `Rgraphviz`.

**Usage**

```r
plotComplex(complexMembers, g, VBs, VPs, geneName = FALSE, baitColor = "yellow", preyColor = "white", recipLineColor = "red", unrecipBBLineColor = "blue", unrecipBPLineColor = "gray", y = "neato")
```

**Arguments**

- `complexMembers`: A character vector of proteins composing a complex estimate.
- `g`: An object of class graph, the full bait-prey graph of AP-MS data used in analysis. `complexMembers` must be a subset of the node names of `g`.
- `VBs`: A vector of viable baits used in the AP-MS experiment.
- `VPs`: A vector of viable prey used in the AP-MS experiment.
- `geneName`: A logical indicating whether or not nodes should be plotted with common gene names as labels rather than systematic names.
- `baitColor`: Color of bait nodes.
- `preyColor`: Color of prey nodes.
- `recipLineColor`: Color of edges connecting baits which both detected each other as prey.
- `unrecipBBLineColor`: Color of edges connecting baits in which one bait finds the other as prey but not vice versa.
- `unrecipBPLineColor`: Color of edges extending from baits to proteins that were only used as prey, hence reciprocity is not possible.
- `y`: Layout of plot

**Details**

This is a simple function for plotting complex estimates resulting from the `apComplex` algorithm. Giving the upcoming changes in `Rgraphviz`, it will likely be changed substantially.

**Value**

A plotted graph of the complex estimate subgraph.
sortComplexes

Author(s)
Denise Scholtens

References

See Also
findComplexes

Examples

data(apEX)
data(apEXG)
PCMG2 <- findComplexes(apEX,sensitivity=.7,specificity=.75)
PCMG2sorted <- sortComplexes(PCMG2,apEX)

VBs <- rownames(apEX)
VPs <- setdiff(colnames(apEX),VBs)

plotComplex(PCMG2sorted$MBME[[1]],g=apEXG,VBs=VBs, VPs=VPs)

sortComplexes Sort complex estimates

Description
Sorts complexes recorded in PCMG list into three separate lists containing MBME, SBMH, and UnRBB complexes.

Usage
sortComplexes(PCMG, adjMat)

Arguments
PCMG Current PCMG estimate
adjMat Adjacency matrix of bait-hit data from an AP-MS experiment. Rows correspond to baits and columns to hits.

Details
MBME complexes contain multiple bait proteins and multiple edges. SBMH complexes contain one bait and a collection of hit-only proteins. UnRBB complexes contain only two baits (no hit-only proteins) that are connected by an unreciprocated edge.
**yNameTAP complexes**

**Value**

A list of lists representing the MBME, SBMH, and UnRBB complex estimates.

**Author(s)**

Denise Scholtens

**References**


**See Also**

findComplexes

**Examples**

```r
data(apEX)
PCMG2 <- findComplexes(apEX,sensitivity=.7,specificity=.75)
sortComplexes(PCMG2,apEX)
```

---

**yNameTAP complexes**

*TAP data complex estimates using standard gene names*

**Description**

Affiliation matrices with rows corresponding to proteins and columns corresponding to complexes.

**Usage**

```r
data(yNameTAP)
```

**Details**

These are the results from an analysis of the TAP data (Gavin et al., 2002) by Scholtens and Gentleman (2004) and Scholtens, Vidal, and Gentleman (submitted). These estimates were constructed using findComplexes with a sensitivity parameter of .75, specificity of .995, and Beta=-0.2 for externally derived similarity measure based on Gene Ontology cellular component annotation (see Scholtens and Gentleman (2004)).

yNameTAP contains 260 multi-bait-multi-edge complex estimates.

**Source**


References


See Also

TAP, TAPgraph, \code{TAPgraph}, \code{findComplexes}

Examples

data(yNameTAP)
yNameTAP[1:10,1:3]

---

yTAP

\code{yTAP Complexes published Gavin, et al. (2002).}

Description

A list of the 232 \code{yTAP} complexes published by Gavin, et al. (2002) using their TAP data.

Usage

data(yTAP)

Details

Each element of the \code{yTAP} list contains a vector of the proteins reported as part of the complex, as well as the functional annotation category published by Gavin, et al. (2002).

These data are available in Supplementary Table S3 of Gavin, et al. (2002) at http://www.nature.com and at http://yeast.cellzome.com.

Source


See Also

TAP, TAPgraph, MBMEcTAP, SBMHcTAP, UnRBBcTAP, HMSPCI

Examples

data(yTAP)
yTAP[1]
Index

*Topic array
- bhmaxSubgraph, 9

*Topic datasets
- apEX, 8
  - gavinBP2006, 12
  - HMSPCI, 2
  - HMSPCI complexes, 5
  - Krogan, 3
  - Krogan complexes, 6
  - kroganBPMat2006, 13
  - TAP, 7
  - TAP complexes, 1, 6
  - yNameTAP complexes, 17
  - yTAP, 18
- LCjoin, 4
  - LCjoinAdjBin (LCjoin), 4
  - LCjoinFisher (LCjoin), 4
  - LCjoinLchange (LCjoin), 4
  - MBMEcHMSPCI, 3
  - MBMEcHMSPCI (HMSPCI complexes), 5
  - MBMEcKrogan, 3
  - MBMEcKrogan (Krogan complexes), 6
  - MBMEcTAP, 8, 12, 13, 18
  - MBMEcTAP (TAP complexes), 6
  - mergeComplexes, 4, 10, 12, 13
  - plotComplex, 15
  - SBMHCcHMSPCI, 3
  - SBMHCcHMSPCI (HMSPCI complexes), 5
  - SBMHCcTAP, 8, 12, 13, 18
  - SBMHCcTAP (TAP complexes), 6
  - sortComplexes, 16
  - TAP, 3, 7, 7, 18
  - TAP complexes, 1, 6
  - TAPgraph, 7, 18
  - TAPgraph (TAP), 7
  - UnRBBcHMSPCI, 3
  - UnRBBcHMSPCI (HMSPCI complexes), 5
  - UnRBBcTAP, 8, 12, 13, 18
  - UnRBBcTAP (TAP complexes), 6
  - yNameTAP (yNameTAP complexes), 17
  - yNameTAP complexes, 17
  - yTAP, 8, 12, 13, 18

*Topic graphs
- findComplexes, 10
  - LCjoin, 4
  - mergeComplexes, 13
  - plotComplex, 15
  - sortComplexes, 16
  - adjBinFUN (LCjoin), 4
    - apEX, 8
    - apEXG (apEX), 8
    - bhmaxSubgraph, 4, 9, 12, 14
  - findComplexes, 4, 10, 10, 14, 16, 17
  - fisherFUN (LCjoin), 4
  - gavin06FilteredEstimates (TAP complexes), 1
  - gavinBP2006, 2, 12
  - gavinBP2006graph (gavinBP2006), 12
  - HMSPCI, 2, 5, 8, 12, 13, 18
  - HMSPCI complexes, 5
  - HMSPCIgraph, 5
  - HMSPCIgraph (HMSPCI), 2
  - Krogan, 3
  - Krogan complexes, 6
  - krogan06FilteredEstimates (TAP complexes), 1
  - kroganBPMat2006, 2, 13