Package 'SMAD'

July 10, 2025

Type Package

Title Statistical Modelling of AP-MS Data (SMAD)

Version 1.24.0

Description Assigning probability scores to protein interactions captured in affinity purification mass spectrometry (AP-MS) expriments to infer protein-protein interactions. The output would facilitate non-specific background removal as contaminants are commonly found in AP-MS data.

Depends R (>= 3.6.0), RcppAlgos

Imports magrittr (>= 1.5), dplyr, stats, tidyr, utils, Rcpp (>= 1.0.0)

LinkingTo Rcpp

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

Suggests knitr, rmarkdown, testthat, BiocStyle

VignetteBuilder knitr

biocViews MassSpectrometry, Proteomics, Software

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

git_branch RELEASE_3_21

git_last_commit a2e2c91

git_last_commit_date 2025-04-15

Repository Bioconductor 3.21

Date/Publication 2025-07-09

Author Qingzhou Zhang [aut, cre]

Maintainer Qingzhou Zhang <zqzneptune@hotmail.com>

Contents

`CompPASS` *CompPASS*

Description

CompPASS Comparative Proteomic Analysis Software Suite (CompPASS) is based on spoke model. This algorithm was developed by Dr. Mathew Sowa for defining the human deubiquitinating enzyme interaction landscape (Sowa, Mathew E., et al., 2009). The implementation of this algorithm was inspired by Dr. Sowa's online tutorial (http://besra.hms.harvard.edu/ipmsmsdbs/cgi-bin/tutorial.cgi). The output includes Z-score, S-score, D-score and WD-score. This function also computes entropy and normalized WD-score. The source code for this function was based on the source code. https://github.com/dnusinow/cRomppass

Usage

```
CompPASS(datInput)
```

Arguments

`datInput`	A dataframe with column names: idRun, idBait, idPrey, countPrey. Each row represent one unique protein captured in one pull-down experiment

Value

A data frame consists of unique bait-prey pairs with Z-score, S-score,D-score and WD-score indicating interacting probabilities.

Author(s)

Qingzhou Zhang, <zqzneptune@hotmail.com>

References

Sowa, Mathew E., et al. "Defining the human deubiquitinating enzyme interaction landscape." Cell 138.2 (2009): 389-403. https://doi.org/10.1016/j.cell.2009.04.042

Huttlin, Edward L., et al. "The BioPlex network: a systematic exploration of the human interactome." Cell 162.2 (2015): 425-440. https://doi.org/10.1016/j.cell.2015.06.043

Huttlin, Edward L., et al. "Architecture of the human interactome defines protein communities and disease networks." Nature 545.7655 (2017): 505. https://www.nature.com/articles/nature22366

Examples

```
data(TestDatInput)
datScore <- CompPASS(TestDatInput)
head(datScore)
```

DICE *DICE*

Description

DICE The Dice coefficient is used to score the interaction affinity between two proteins.

Usage

```
DICE(datInput)
```

Arguments

datInput	A dataframe with column names: idRun, idPrey. Each row represent one unique protein captured in one pull-down experiment.

Value

A dataframe consists of pairwise combindation of preys identified in the input with DICE scores.

Author(s)

Qingzhou Zhang, <zqzneptune@hotmail.com>

References

Bing Zhang et al., From pull-down data to protein interaction networks and complexes with biological relevance, Bioinformatics, Volume 24, Issue 7, 1 April 2008, Pages 979–986, https://doi.org/10.1093/bioinformatics/btn036

Examples

```
data(TestDatInput)
datScore <- DICE(TestDatInput)
head(datScore)
```

Hart *Hart*

Description

Hart Scoring algorithm based on a hypergeometric distribution error model (Hart et al.,2007).

Usage

```
Hart(datInput)
```

Arguments

datInput	A dataframe with column names: idRun, idPrey. Each row represent one unique protein captured in one pull-down experiment.

Value

A dataframe consists of pairwise combindation of preys identified in the input with Hart scores indicating interacting probabilities computed from negative log transformed Hypergeometric test P-values.

Author(s)

Qingzhou Zhang, <zqzneptune@hotmail.com>

References

Hart, G. Traver, Insuk Lee, and Edward M. Marcotte. 'A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality.' BMC bioinformatics 8.1 (2007): 236. https://doi.org/10.1186/1471-2105-8-236

Examples

```
data(TestDatInput)
datScore <- Hart(TestDatInput)
head(datScore)
```

HG *HGScore*

Description

HGScore Scoring algorithm based on a hypergeometric distribution error model (Hart et al.,2007) with incorporation of NSAF (Zybailov, Boris, et al., 2006) . This algorithm was first introduced to predict the protein complex network of Drosophila melanogaster (Guruharsha, K. G., et al., 2011). This scoring algorithm was based on matrix model.

Usage

```
HG(datInput)
```

Arguments

`datInput`	A dataframe with column names: idRun, idPrey, countPrey, lenPrey. Each row represent one unique protein captured in one pull-down experiment.

Value

A dataframe consists of pairwise combindation of preys identified in the input with HG scores indicating interacting probabilities computed from negative log transformed Hypergeometric test P-values.

Author(s)

Qingzhou Zhang, <zqzneptune@hotmail.com>

References

Guruharsha, K. G., et al. 'A protein complex network of Drosophila melanogaster.' Cell 147.3 (2011): 690-703. https://doi.org/10.1016/j.cell.2011.08.047

Hart, G. Traver, Insuk Lee, and Edward M. Marcotte. 'A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality.' BMC bioinformatics 8.1 (2007): 236. https://doi.org/10.1186/1471-2105-8-236

Zybailov, Boris, et al. 'Statistical analysis of membrane proteome expression changes in Saccharomyces c erevisiae.' Journal of proteome research 5.9 (2006): 2339-2347. https://doi.org/10.1021/pr060161n

Examples

```
data(TestDatInput)
datScore <- HG(TestDatInput)
head(datScore)
```

PE *PE*

Description

PE Incorporated both spoke and matrix model.

Usage

```
PE(datInput, rBait = 0.37, cntPseudo = 1)
```

Arguments

datInput	A dataframe with column names: idRun, idBait, idPrey, Each row represent one unique protein captured in one pull-down experiment.
rBait	The value of the 'r' parameter as desribed in the publication.
cntPseudo	The value of the 'pseudo count' parameter.

Value

A dataframe consists of protein-protein interactions from both Spoke and Matrix model.

Author(s)

Qingzhou Zhang, <zqzneptune@hotmail.com>

References

Collins, Sean R., et al. "Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae." Molecular & Cellular Proteomics 6.3 (2007): 439-450. https://doi.org/10.1074/mcp.M600381-MCP200

Examples

```
data(TestDatInput)
datScore <- PE(TestDatInput, 0.37, 1)
head(datScore)
```

TestDatInput *Test data for SMAD*

Description

It is a subset of unfiltered BioPlex 2.0 consisting of apoptosis as bait proteins

Usage

```
data(TestDatInput)
```

Format

A data frame with 5000 rows and 5 variables

Details

- idRun A unique identifier for one AP-MS run
- idBait A unique identifier for the bait protein
- idPrey A unique identifier for the prey protein
- countPrey Sepctra/Peptider count for the prey protein
- lenPrey Protein length for the prey protein

Index