# Package 'MSstatsBioData'

April 11, 2019

**Type** Package

**Title** Datasets of published biological studies with DDA or SRM experiments

**Version** 1.4.0

**Date** 2017-10-02

**Author** Meena Choi

**Maintainer** Meena Choi <mnchoi67@gmail.com>

**Description** Provides the peak intensity data for detecting differentially abundant proteins in seven published biological investigations.

**License** Artistic-2.0

**Depends** R (>= 3.4)

**Suggests** BiocStyle, knitr, MSstats

**VignetteBuilder** knitr

**biocViews** ExperimentData, MassSpectrometryData, Proteome

**NeedsCompilation** no

**Encoding** UTF-8

**LazyData** true

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

**git_branch** RELEASE_3_8

**git_last_commit** ae30d7b

**git_last_commit_date** 2018-10-30

**Date/Publication** 2019-04-11

## R topics documented:

`MSstatsBioData-package` *Datasets of published biological studies with DDA or SRM experiments*

## Description

Provides the peak intensity data for detecting differentially abundant proteins between groups. Seven datasets from published biological investigations are available. (see `DDA_cardio`, `SRM_yeast`, `SRM_ovarian`, `SRM_crc_training`, `SRM_crc_validation`, `SRM_mpm_training`, `SRM_mpm_validation`)

## Details

All datasets was processed as described in original reference. They were reformatted as MSstats required format.

To view the example workflows, type `browseVignettes("MSstatsBioData")`.

## Author(s)

Meena Choi

Maintainer: Meena Choi <mnchoi67@gmail.com>

## References

Clough, T. et al. (2009) Protein quantification in label-free LC-MS experiments. *J. Proteome Res.*, 8, 5275–5284.

Picotti, P. et ak. (2009) Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell*, 138, 795–806.

Huttenhain, R. et al. (2012) Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Sci. Tansl. Med.*, 4, 142ra94.

Cerciello, F. et al. (2013) Identification of a seven glycopeptide signature for malignant pleural mesothelioma in human serum by selected reaction monitoring. *Clin. Proteomics*, 10, 16.

Surinova, S. et al. (2015) Prediction of colorectal cancer diagnosis based on circulating plasma proteins. *EMBO Mol. Med.*, 7, 1166–1178.

`DDA_cardio` *Dataset of cardiovascular disease study*

## Description

This study is for investigation for cardiovascular disease between control and four disease stages. (0, 1, 2, 3, 4 in Condition) 246 samples from control and disease patients were analyzed with single injection by label- free DDA as described in 15. There are 97 identified proteins. The dataset was processed by Monarch, (http: //www.bloomberg.com/research/stocks/private/snapshot.asp?privcapid=20704167). Unusually, this DDA dataset had no missing values because the procedure reported the background signal if a feature in a run was not detected.

## Usage

```
data(DDA_cardio)
```

## Format

DDA_cardio is a data.frame.

## References

Clough, T. et al. (2009) Protein quantification in label-free LC-MS experiments. *J. Proteome Res.*, 8, 5275–5284.

## Examples

```
data(DDA_cardio)

## Example of using MSstats for differential abundance analysis.
require(MSstats)
input.proposed <- dataProcess(DDA_cardio,
                              summaryMethod="TMP",
                              cutoffCensored="minFeature",
                              censoredInt="NA",
                              MBimpute=TRUE,
                              maxQuantileforCensored=0.999)

comparison1<-matrix(c(-1,1,0,0,0),nrow=1)
comparison2<-matrix(c(-1,0,1,0,0),nrow=1)
comparison3<-matrix(c(-1,0,0,1,0),nrow=1)
comparison4<-matrix(c(-1,0,0,0,1),nrow=1)

comparison<-rbind(comparison1, comparison2, comparison3, comparison4)
row.names(comparison)<-c("1-0","2-0","3-0","4-0")

output.comparison <- groupComparison(contrast.matrix=comparison,
                              data=input.proposed)
head(output.comparison$ComparisonResult)
```

---

# SRM_crc_training *The training set from a study for subjects with colorectal cancer*

---

## Description

72 proteins, including two standard proteins, AIAG-Bovine and FETUA-Bovine, were targeted for plasma samples with SRM with isotope labeled reference peptides in order to identify candidate protein biomarker for non-invasive detection of CRC. The training cohort included 100 subjects in control group and 100 subjects with CRC. Each sample for subject was measured in a single injection without technical replicate. The training cohort was analyzed with Skyline. The dataset was already normalized as described in manuscript. User do not need extra normalization. NAs should be considered as censored missing. Two standard proteins can be removed for statistical analysis.

## Usage

```
data(SRM_crc_training)
```

## Format

SRM_crc_training is a data.frame.

## References

Surinova, S. et al. (2015) Prediction of colorectal cancer diagnosis based on circulating plasma proteins. *EMBO Mol. Med.*, 7, 1166–1178.

## Examples

```
## Intensities are already normalized as described in the reference.
data(SRM_crc_training)

## Example of using MSstats for differential abundance analysis.
require(MSstats)
input.proposed <- dataProcess(SRM_crc_training,
                              normalization=FALSE,
                              summaryMethod="TMP",
                              cutoffCensored="minFeature",
                              censoredInt="NA",
                              MBimpute=TRUE,
                              maxQuantileforCensored=0.999)

comparison<-matrix(c(1,-1),nrow=1)
row.names(comparison)<-c("Disease-Healthy")

output.comparison <- groupComparison(contrast.matrix=comparison,
                              data=input.proposed)
head(output.comparison$ComparisonResult)
```

---

# SRM_crc_validation *The validation set from a study for subjects with colorectal cancer*

---

## Description

72 proteins, including two standard proteins, AIAG-Bovine and FETUA-Bovine, were targeted for plasma samples with SRM with isotope labeled reference peptides in order to identify candidate protein biomarker for non-invasive detection of CRC. The validation cohort had 67 subjects in controls, and 202 subject with different clinical stages of CRC. Each sample for subject was measured in a single injection without technical replicate. The validation cohort was processed with MultiQuant 1.2. NAs should be considered as censored missing. Two standard proteins can be removed for statistical analysis.

## Usage

```
data(SRM_crc_validation)
```

## Format

SRM_crc_validation is a data.frame.

## References

Surinova, S. et al. (2015) Prediction of colorectal cancer diagnosis based on circulating plasma proteins. *EMBO Mol. Med.*, 7, 1166–1178.

## Examples

```
data(SRM_crc_validation)

## Example of using MSstats for differential abundance analysis.
require(MSstats)
input.proposed <- dataProcess(SRM_crc_validation,
                              normalization=FALSE,
                              summaryMethod="TMP",
                              cutoffCensored="minFeature",
                              censoredInt="NA",
                              MBimpute=TRUE,
                              maxQuantileforCensored=0.999)

comparison<-matrix(c(1,-1),nrow=1)
row.names(comparison)<-c("Disease-Healthy")

output.comparison <- groupComparison(contrast.matrix=comparison,
                              data=input.proposed)
head(output.comparison$ComparisonResult)
```

---

# SRM_mpm_training *The training set from a study of subjects with malignant pleural mesothelioma(MPM)*

---

## Description

To identify candidate biomarkers for MPM in serum, the experiment targeted 32 candidate peptides with SRM with isotope labeled reference peptides. For peptide-level analysis, ProteinName column has unique id for each peptide. The training set includes total 75 subjects: 25 MPM, 25 healthy donors(HD), 25 non-small cell lung cancer (NSCLC). Each sample was injected once without technical replicate. All samples were processed by Skyline. Zero value in Intensity should be considered as censored missing.

## Usage

```
data(SRM_mpm_training)
```

## Format

SRM_mpm_training is a data.frame.

## References

Cerciello, F. et al. (2013) Identification of a seven glycopeptide signature for malignant pleural mesothelioma in human serum by selected reaction monitoring. *Clin. Proteomics*, 10, 16.

## Examples

```
data(SRM_mpm_training)

## Example of using MSstats for differential abundance analysis.
require(MSstats)
input.proposed <- dataProcess(SRM_mpm_training,
                              summaryMethod="TMP",
                              cutoffCensored="minFeature",
                              censoredInt="0",
                              MBimpute=TRUE,
                              maxQuantileforCensored=0.999)

comparison1<-matrix(c(-1,1,0),nrow=1)
comparison2<-matrix(c(-1,0,1),nrow=1)
comparison3<-matrix(c(0,1,-1),nrow=1)
comparison<-rbind(comparison1, comparison2, comparison3)
row.names(comparison)<-c("MPM-control","NSCLC-control","MPM-NSCLC")

output.comparison <- groupComparison(contrast.matrix=comparison,
                              data=input.proposed)
head(output.comparison$ComparisonResult)
```

---

`SRM_mpm_validation` *The validation set from a study of subjects with malignant pleural mesothelioma(MPM)*

---

## Description

To identify candidate biomarkers for MPM in serum, the experiment targeted 31 candidate peptides with SRM with isotope labeled reference peptides. The validation set consists of total 98 subjects: 34 MPM, 32 healthy donors(HD), 32 non-small cell lung cancer (NSCLC). Each sample was injected once without technical replicate. 7 Subjects are overlapped with training set. All samples were processed by Skyline.

## Usage

```
data(SRM_mpm_validation)
```

## Format

`SRM_mpm_validation` is a data.frame.

## References

Cerciello, F. et al. (2013) Identification of a seven glycopeptide signature for malignant pleural mesothelioma in human serum by selected reaction monitoring. *Clin. Proteomics*, 10, 16.

## Examples

```
data(SRM_mpm_validation)

## Example of using MSstats for differential abundance analysis.
require(MSstats)
input.proposed <- dataProcess(SRM_mpm_validation,
                              summaryMethod="TMP",
                              cutoffCensored="minFeature",
                              censoredInt="0",
                              MBimpute=TRUE,
                              maxQuantileforCensored=0.999)

comparison1<-matrix(c(-1,1,0),nrow=1)
comparison2<-matrix(c(-1,0,1),nrow=1)
comparison3<-matrix(c(0,1,-1),nrow=1)
comparison<-rbind(comparison1, comparison2, comparison3)
row.names(comparison)<-c("MPM-control","NSCLC-control","MPM-NSCLC")

output.comparison <- groupComparison(contrast.matrix=comparison,
                              data=input.proposed)
head(output.comparison$ComparisonResult)
```

---

# SRM_ovarian *Dataset for a study of subjects with ovarian cancer*

---

## Description

Original published raw data, SRM with isotope labeled reference peptides, has total 83 patients plasma samples. Skyline succeeded to analyze 81 patients samples. The dataset including 66 ovarian cancer (OC) patients and 15 patients with benign ovarian tumors was used to evaluate. Each patient sample measured once without technical replicate. Total 36 proteins were used to evaluate the ability of statistical method to detect differential abundance proteins between OC and benign groups.

## Usage

```
data(SRM_ovarian)
```

## Format

SRM_ovarian is a data.frame.

## References

Huttenhain, R. et al. (2012) Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Sci. Tansl. Med.*, 4, 142ra94.

## Examples

```
data(SRM_ovarian)

## Example of using MSstats for differential abundance analysis.
require(MSstats)
input.proposed <- dataProcess(SRM_ovarian,
                              summaryMethod="TMP",
                              cutoffCensored="minFeature",
                              censoredInt="0",
                              MBimpute=TRUE,
                              maxQuantileforCensored=0.999)

comparison<-matrix(c(1,-1),nrow=1)
row.names(comparison)<-c("Disease-Healthy")

output.comparison <- groupComparison(contrast.matrix=comparison,
                              data=input.proposed)
head(output.comparison$ComparisonResult)
```

---

# SRM_yeast — *Time course investigation of central carbon metabolism of* S. cerevisiae

---

## Description

45 proteins in the glycolysis/gluconeogenesis/TCA cycle/glyoxylate cycle network were targeted in SRM experiment with isotope labeled reference peptides. Three biological replicates were measured at ten time points (T1-T10, labeled as 1 to 10 in Condition column). There are total 30 MS runs measured. It covered dynamic growth phases of *S. cerevisiae*, in a glucose-rich medium (T1-T4), diauxic shift (T5-T6), post-diauxic phase (T7-T9), and stationary phase (T10). Each transition was quantified automatically using MultiQuant with no missing values.

## Usage

```
data(SRM_yeast)
```

## Format

`SRM_yeast` is a data.frame.

## References

Picotti, P. et ak. (2009) Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell*, 138, 795–806.

## Examples

```
data(SRM_yeast)

## Example of using MSstats for differential abundance analysis.
require(MSstats)
input.proposed <- dataProcess(SRM_yeast,
```

```
                            summaryMethod="TMP",
                            cutoffCensored="minFeature",
                            censoredInt="0",
                            MBimpute=TRUE,
                            maxQuantileforCensored=0.999)

comparison1<-matrix(c(-1,1,0,0,0,0,0,0,0,0),nrow=1)
comparison2<-matrix(c(-1,0,1,0,0,0,0,0,0,0),nrow=1)
comparison3<-matrix(c(-1,0,0,1,0,0,0,0,0,0),nrow=1)
comparison4<-matrix(c(-1,0,0,0,1,0,0,0,0,0),nrow=1)
comparison5<-matrix(c(-1,0,0,0,0,1,0,0,0,0),nrow=1)
comparison6<-matrix(c(-1,0,0,0,0,0,1,0,0,0),nrow=1)
comparison7<-matrix(c(-1,0,0,0,0,0,0,1,0,0),nrow=1)
comparison8<-matrix(c(-1,0,0,0,0,0,0,0,1,0),nrow=1)
comparison9<-matrix(c(-1,0,0,0,0,0,0,0,0,1),nrow=1)

comparison <- rbind(comparison1,comparison2,comparison3,
                    comparison4,comparison5,comparison6,
                    comparison7,comparison8,comparison9)
row.names(comparison) <- c("T2-T1","T3-T1","T4-T1",
                           "T5-T1","T6-T1","T7-T1",
                           "T8-T1","T9-T1","T10-T1")


output.comparison <- groupComparison(contrast.matrix=comparison,
                           data=input.proposed)
head(output.comparison$ComparisonResult)
```

# Index