Package ‘proteomics’

February 20, 2015

Version 0.2
Date 2014-11
Title Statistical Analysis of High Throughput Proteomics Data
Description Provides methods for making inference in isobaric labelled LC-MS/MS experiments, i.e. iTRAQ experiments. It provides a function that reasonably parses a CSV-export from Proteome Discoverer(TM) into a data frame that can be easily handled in R. Functions and methods are provided for quality control, filtering, norming, and the calculation of response variables for further analysis. The merging of multiple iTRAQ experiments with respect to a reference is also covered.

Author Thomas W. D. Möbius <kontakt@thomasmoebius.de>
Maintainer Thomas W. D. Möbius <kontakt@thomasmoebius.de>
URL http://00tau.github.io/proteomics-in-r/
License GPL-3
Imports plyr, reshape2, ggplot2, foreach
LazyData yes
ByteCompile yes
Encoding UTF-8
Collate 'data-00-selection-from-PD.r' 'data-01-consolidation.r'
   'data-02-merging.r' 'data-03-analysis.r' 'zzz.R'
NeedsCompilation no
Repository CRAN
Date/Publication 2014-11-22 01:30:39

R topics documented:

  accum ......................................................... 2
  addIonStatistics ............................................ 3
  addLoadings ................................................. 3
  addRetentionAtApex ......................................... 4
addRetentionIndexTimeStatistics ........................................ 4
adjustBy ........................................................................... 5
adjusting ......................................................................... 5
adjustOne .......................................................................... 6
avrgLoading ................................................................. 6
channelResponses ......................................................... 6
copyLoadings ............................................................... 7
factoring .......................................................................... 7
meetSelection ..................................................................... 8
mergeFrames ..................................................................... 9
norm2Reference .............................................................. 9
pAction ............................................................................ 10
plotMePeptide .................................................................. 10
plotMeProtein ................................................................... 10
pRetention ......................................................................... 11
pVioline ........................................................................... 11
pVolcano .......................................................................... 12
responseStatistics .......................................................... 12
selectByConfidence ....................................................... 13
selectByEffect ............................................................... 13
selectByFDR ..................................................................... 14
testForPeptideEffect .................................................... 14
testForProteinEffect ..................................................... 15
testing .............................................................................. 15
testingOneshot ............................................................. 16
testingTukey ................................................................. 16
toAlpha ............................................................................ 16
toProportions ............................................................... 17

Index 18

accum Response calculation

Description
Calculates needed sample size accumulation from iTRAQ data which is given on spectrum level.

Usage
accum(dwide)

Arguments
dwide iTRAQ data in wide format including columns corresponding to iTRAQ channels containing their intensities.
addIonStatistics

Summary statistics – Ion intensities per spectra

Description

Summary statistics – Ion intensities per spectra

Usage

addIonStatistics(dwide)

Arguments

dwide iTRAQ data in wide format including columns corresponding to iTRAQ channels containing their intensities.

addLoadings

Adjust for confounding – add an appropriate target

Description

Adjust for confounding – add an appropriate target

Usage

addLoadings(dwide, byRef = F)

Arguments

dwide iTRAQ data in wide format

byRef shouold the average be calculated from the loading of the reference channel. Default is FALSE and this is recommended.
### addRetentionAtApex

*Summary statistics – Calculates retention time statistics at apex*

#### Description
Calculates different summary retention time statistics for each peptide (a subsequence of a protein including post translational modifications). The idea is that each peptide is supposed to have roughly the same retention time.

#### Usage
```
addRetentionAtApex(dwide, ...)
```

#### Arguments
- `dwide` : iTRAQ data in wide format
- `...` : Additional arguments passed for `ddply`

### addRetentionIndexTimeStatistics

*Summary statistics – Calculates index retention time statistics*

#### Description
Summary statistics – Calculates index retention time statistics

#### Usage
```
addRetentionIndexTimeStatistics(dwide, ...)
```

#### Arguments
- `dwide` : iTRAQ data in wide format
- `...` : Additional arguments passed for `ddply`
**adjustBy**

*Adjust for confounding – Generic function for centring data*

**Description**

This function calculates from given adjusting factors that compensate for possible confounding due the transformed values for the statistical analysis.

**Usage**

```r
adjustBy(dwide, effect, ch)
```

**Arguments**

- `dwide`: iTRAQ data in wide format.
- `effect`: estimated effects which may yield to confounding.
- `ch`: names of the channel columns.

**Details**

Can be used to perform custom adjustments. (Code not used anymore.)

**adjusting**

*Adjust for confounding – State of the art adjustments for confounding*

**Description**

Compensate for possible confounding due the transformed values for the statistical analysis.

**Usage**

```r
adjusting(dwide)
```

**Arguments**

- `dwide`: iTRAQ data in wide format.
adjustOne

Adjust for confounding – In one single experiment only

Description
Simple code when only one iTRAQ-experiment has been performed. (Code not used anymore.)

Usage
adjustOne(dwide)

Arguments
- dwide: iTRAQ data in wide format.

avrgLoading

Adjust for confounding – calculates the average loading

Description
Adjust for confounding – calculates the average loading

Usage
avrgLoading(dwide)

Arguments
- dwide: iTRAQ data in wide format

channelResponses

Response calculation

Description
From spectrum to protein level – Response variable calculation

Usage
channelResponses(dwide, acc)

Arguments
- dwide: iTRAQ data in wide format including columns corresponding to iTRAQ channels containing their intensities.
- acc: result of an accumulation of sample sizes
Description

This is important when analysing enriched samples. Here, use the loading averages from the corresponding non-enriched sample.

Usage

```r
copyLoadings(fromWide, toWide)
```

Arguments

- `fromWide`: iTRAQ data in wide format
- `toWide`: iTRAQ data in wide format

Description

Making a multiple-factor ANOVA from the single channel variable of an iTRAQ experiment.

Usage

```r
factoring(dwide, cvmat)
```

Arguments

- `dwide`: iTRAQ data in wide format including columns corresponding to iTRAQ channels containing their intensities.
- `cvmat`: a matrix that hold the information on which channel is mapped to which factor.

Details

This function uses a matrix `cvmat` to convert the single channel into a full fledged multiple factor ANOVA.
**Examples**

```r
channels <- c("X113", "X114", "X115", "X116", "X117", "X118", "X119") #, "X121")
typus <- c(rep(c("A", "B", "C"), each=2), "reference")
treatment <- c(rep(c("I", "II"), 3), "mixed")
convmat <- data.frame(channels=channels, typus=typus, treatment=treatment)
print(convmat)
## Not run: factoring(dwide, cvmat=convmat)
```

**Description**

Has been tested with PD v1.4

**Usage**

```r
meetSelection(dwide, ch, ref)
```

**Arguments**

- **dwide**: raw data from a PD export.
- **ch**: the column names which hold the reporter ion intensities.
- **ref**: the column name which holds the reporter ion intensities of the reference channel.

**Details**

This is a rather neat function that allows to get data from an export form the software Proteom Discoverer into R and parsed into a reasonable data frame such one can work with it. It will also add a few statistics and create unique identifiers for all identified peptides. You may argue that this functionality alone is worth the import of the whole package.

**Examples**

```r
bioQ <- read.csv("my-proteome-discoverer-v1.4-export-experiment-1.csv")
bioR <- read.csv("my-proteome-discoverer-v1.4-export-experiment-2.csv")
run1 <- droplevels(bioQ[bioQ$Quan.Usage == "Used",])
run2 <- droplevels(bio2[bio2$Quan.Usage == "Used",])
channels <- c("X113", "X114", "X115", "X116", "X117", "X118", "X119", "X121")
reference <- c("X121")
run1 <- meetSelection(run1, channels, reference)
run2 <- meetSelection(run2, channels, reference)
run1$experiment <- factor(1, levels=1:2, labels=c("iTRAQ-1", "iTRAQ-2"))
run2$experiment <- factor(2, levels=1:2, labels=c("iTRAQ-1", "iTRAQ-2"))
rungs <- rbind(run1, run2)
## End(Not run)
```
mergeFrames

Merging multiple experiments

Description

At the end each channel in each iTRAQ experiment can be uniquely identified by a barcode. If two channels of different experiments correspond to the same subject, the same barcode may be used and a method of combining these measurements be chosen.

Usage

mergeFrames(files, path, sampledesign)

Arguments

files data frame of file names and corresponding ids.
path leading to the files
sampledesign data frame of ids, channel names and corresponding barcodes.

norm2Reference Response calculation

Description

Norming the responses of a single iTRAQ to a given reference channel.

Usage

norm2Reference(dlong)

Arguments

dlong iTRAQ data in long format.
### pAction

*Plotting p-value distributions*

**Description**
Plotting p-value distributions

**Usage**

```r
pAction(restest)
```

**Arguments**

- `restest` : result frame of test results

### plotMePeptide

*Plot interaction plots of peptides*

**Description**
Plot interaction plots of peptides

**Usage**

```r
plotMePeptide(datP)
```

**Arguments**

- `datP` : subframe of peptide data

### plotMeProtein

*Plot interaction plots of proteins*

**Description**
Plot interaction plots of proteins

**Usage**

```r
plotMeProtein(datP)
```

**Arguments**

- `datP` : subframe of protein data
**pRetention**

*Plot Retention Time Statistics*

**Description**

Plot retention times with possible outliers

**Usage**

```r
pRetention(rwide)
```

**Arguments**

- `rwide` iTRAQ data in wide format with retention time information

**Examples**

```r
## Not run:
iglobal <- addIonStatistics(pglobal)
rglobal <- addRetentionTimeStatistics(iglobal, .parallel=TRUE)
rglobal$outlier <- with(rglobal, abs(retention.atApex - retention) > 4)
p <- pRetention(rglobal)

p + geom_point(aes(retention.atApex, retention))
p + geom_point(aes(retention.atApex, retention-retention.atApex))
p + geom_point(aes(ppm, retention-retention.atApex))
p + geom_density(aes(x=ppm), alpha=.242)

## End(Not run)
```

---

**pVioline**

*Plot Retention Time Statistics in violine form*

**Description**

Plot Retention Time Statistics in violine form

**Usage**

```r
pVioline(dat, target)
```

**Arguments**

- `dat` iTRAQ in log format
- `target` of the norming
pVolcano

Volcano plot

Description
Volcano plot

Usage
pVolcano(res, threshold, .foldchange = TRUE, 
.plot = TRUE)

Arguments
res
result frame of test results
threshold
for biological reasonable effect
.foldchange
whether results given in ratios or log-ratios
.plot
if true adds a plotting layer

responseStatsics
Summary statistics – Generic to calculate summary statistics

Description
Calculates generic summary statistics based on a given formula.

Usage
responseStatsics(dwide, frm)

Arguments
dwide
iTRAQ data in wide format
frm
for example: frm <- value ~ protein + variable frm <- value ~ peptide + variable
selectByConfidence

Result filtering – Test for biological effect

Description

The result file filtered by contains on the confidence intervals. This function will use these confidence intervals to filter out biological irrelevant effects.

Usage

selectByConfidence(res, threshold, foldchange = TRUE)

Arguments

res Result file
threshold Biologically reasonable threshold
foldchange Is the threshold given a fold change or a log2-fold change. Default ist TRUE.

selectByEffect

Result filtering – Test for biological effect

Description

Result filtering – Test for biological effect

Usage

selectByEffect(res, cutoff = 1)

Arguments

res Result file

cutoff the cutoff to be used in the selection
**selectByFDR**  
*Result filtering*

**Description**

Result filtering

**Usage**

```r
selectByFDR(res, fdr = 0.01)
```

**Arguments**

- `res`: result frame of test results
- `fdr`: false discovery rate

**testForPeptideEffect**  
*Data Analysis – Testing on peptide level*

**Description**

Data Analysis – Testing on peptide level

**Usage**

```r
testForPeptideEffect(dat, frm, conf.level, ...)
```

**Arguments**

- `dat`: iTRAQ data in long format
- `frm`: formal for the test
- `conf.level`: confidence level
- `...`: arguments understood by `ddply`
**testForProteinEffect**  
*Data Analysis – Testing on protein level*

**Description**

Data Analysis – Testing on protein level

**Usage**

```
testForProteinEffect(dat, frm, conf.level, ...)
```

**Arguments**

dat  
iTRAQ data in long format
frm  
formal for the test
conf.level  
confidence level
...  
arguments understood by dplyr

---

**testing**  
*Data Analysis – Testing features with Tukey Honest Significant Differences*

**Description**

Data Analysis – Testing features with Tukey Honest Significant Differences

**Usage**

```
testing(dp, frm, conf.level)
```

**Arguments**

dp  
iTRAQ data in long format
frm  
formal for the test
conf.level  
confidence level
testingOneshot  
*Data Analysis – Testing one feature without Tukey Honest Significant Differences*

**Description**
Data Analysis – Testing one feature without Tukey Honest Significant Differences

**Usage**
```
testingOneshot(model)
```

**Arguments**
- `model` : ANOVA model of the corresponding fit

---

testingTukey  
*Data Analysis – Testing one feature with Tukey Honest Significant Differences*

**Description**
Data Analysis – Testing one feature with Tukey Honest Significant Differences

**Usage**
```
testingTukey(model, conf.level)
```

**Arguments**
- `model` : ANOVA model of the corresponding fit
- `conf.level` : confidence level

---

toAlpha  
*Measuring stability – angle of loading vector*

**Description**
Measuring stability by evaluating angle of loading vector from identity

**Usage**
```
toAlpha(dwide)
```

**Arguments**
- `dwide` : iTRAQ data in wide format
toProportions

Transformation – From intensity scales to density histograms

Description

Transformation – From intensity scales to density histograms

Usage

toProportions(dwide)

Arguments

dwide iTRAQ data in wide format
Index

accum, 2
addIonStatistics, 3
addLoadings, 3
addRetentionAtApex, 4
addRetentionIndexTimeStatistics, 4
adjustBy, 5
adjusting, 5
adjustOne, 6
avrgLoading, 6

channelResponses, 6
copyLoadings, 7

factoring, 7

meetSelection, 8
mergeFrames, 9

norm2Reference, 9

pAction, 10
plotMePeptide, 10
plotMeProtein, 10
pRetention, 11
pVioline, 11
pVolcano, 12

responseStatistics, 12

selectByConfidence, 13
selectByEffect, 13
selectByFDR, 14

testForPeptideEffect, 14
testForProteinEffect, 15
testing, 15
testingOneshot, 16
testingTukey, 16
toAlpha, 16
toProportions, 17