Package: mspms (via r-universe)

November 26, 2024

Type Package

Title Tools for the analysis of MSP-MS data

Version 0.99.6

Description This package provides functions for the analysis of data generated by the multiplex substrate profiling by mass spectrometry for proteases (MSP-MS) method. Data exported from upstream proteomics software is accepted as input and subsequently processed for analysis. Tools for statistical analysis, visualization, and interpretation of the data are provided.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

Depends R (>= 4.4.0)

biocViews Proteomics, MassSpectrometry, Preprocessing

LazyData true

Imports QFeatures, SummarizedExperiment, magrittr, rlang, dplyr, purrr, stats, tidyr, stringr, ggplot2, ggseqlogo, heatmaply, readr, rstatix, tibble, ggpubr

Suggests knitr, testthat (>= 3.0.0), downloadthis, DT, rmarkdown, BiocStyle, imputeLCMD

Config/testthat/edition 3

URL https://github.com/baynec2/mspms

BugReports https://github.com/baynec2/mspms/issues

VignetteBuilder knitr

Config/pak/sysreqs cmake libglpk-dev make libmagick++-dev gsfonts libicu-dev libxml2-dev libssl-dev libx11-dev

Repository https://bioc.r-universe.dev

RemoteUrl https://github.com/bioc/mspms

RemoteRef HEAD

RemoteSha 4b8fe4e785d77b38d2a3e008e4c2c5e715d82279

Contents

`all_possible_8mers_from_228_library`

all_possible_8mers_from_228_library All possible 8mers from the standard (as of 26April2024) 228 MSP-MS peptide library (This is equivalent to the result of mspms::calculate_all_cleavages(mspms::peptide_library$real_cleavage_seq,n=4)) vector of the 14 AA peptides used in the library.

Description

all_possible_8mers_from_228_library All possible 8mers from the standard (as of 26April2024) 228 MSP-MS peptide library (This is equivalent to the result of mspms::calculate_all_cleavages(mspms::peptide_library$real vector of the 14 AA peptides used in the library.

Usage

```
all_possible_8mers_from_228_library
```

Format

'all_possible_8mers_from_228_library' A vector with 2964 entries

Source

<standard peptide library used with MSP-MS method in the O'Donoghue lab as of 26April2024>

`calculate_all_cleavages`

calculate_all_cleavages calculate all possible cleavages for a defined peptide library containing peptides of the same length.

Description

calculate_all_cleavages calculate all possible cleavages for a defined peptide library containing peptides of the same length.

Usage

```
calculate_all_cleavages(peptide_library_seqs, n_AA_after_cleavage = 4)
```

Arguments

`peptide_library_seqs`
: The sequences of each peptide in the peptide library. They should all be the same length.

`n_AA_after_cleavage`
: The number of AA after (and before) the cleavage site to consider.

Value

a vector of all the possible cleavages for the peptide library sequences

Examples

```
calculate_all_cleavages(mspms::peptide_library$library_real_sequence,
  n_AA_after_cleavage = 4
)
```

check_file_is_valid_fragpipe

check_file_is_valid_fragpipe Check to make sure the input data looks like the expected FragPipe file.

Description

check_file_is_valid_fragpipe Check to make sure the input data looks like the expected FragPipe file.

Usage

```
check_file_is_valid_fragpipe(fragpipe_data)
```

Arguments

fragpipe_data	combined_peptide.tsv file generated by FragPipe read into R.

Value

a stop command with a informative message if file looks unexpected. otherwise, nothing.

check_file_is_valid_pd

check_file_is_valid_pd Check to make sure the input data looks like the expected ProteomeDiscoverer file.

Description

check_file_is_valid_pd Check to make sure the input data looks like the expected ProteomeDiscoverer file.

Usage

```
check_file_is_valid_pd(pd_data)
```

Arguments

pd_data	PeptideGroups.txt file generated by ProteomeDiscover and read into R.

Value

a stop command with a informative message if file looks unexpected. otherwise, nothing.

`colData` *colData A tibble containing the colData associated with an experiment to proc*

Description

colData A tibble containing the colData associated with an experiment to proc

Usage

```
colData
```

Format

'colData' A tibble: 42 × 4

Source

colData corresponding to cathepsin A-D MSP-MS experiment

`generate_report` *generate_report*

Description

wrapper function to generate an automatic .html report of a basic mspms analysis.

Usage

```
generate_report(
  prepared_data,
  peptide_library = mspms::peptide_library,
  n_residues = 4,
  outdir = getwd(),
  output_file = paste0(Sys.Date(), "_mspms_report.html")
)
```

Arguments

`prepared_data`	a QFeatures object containing a SummarizedExperiment named "peptides".
`peptide_library`	peptide library used with experiment. Contains columns "library_id", "library_match_sequence", and "library_real_sequence".
`n_residues`	the number of amino acid residues before and after the cleavage site to generate a cleavage seq for.
`outdir`	the output directory you would like to render the report to.
`output_file`	the file name to export.

Value

a knited .html report of the mspms analysis.

Examples

```
generate_report(mspms::peaks_prepared_data)
```

`log2fc_t_test` *log2fc_t_test*

Description

Calculates the log2 fold change and t-test statistics given a user specified reference variable and value.

Usage

```
log2fc_t_test(processed_qf, reference_variable = "time", reference_value = 0)
```

Arguments

`processed_qf`	mspms data in a QFeatures object.
`reference_variable`	the colData variable to use as reference
`reference_value`	the value of the colData variable to use as reference

Value

a tibble containing log2fc and t test statistics

Examples

```
log2fc_and_t_test <- log2fc_t_test(mspms::processed_qf)
```

`log2fc_t_test_data` *log2fc_t_test_data A tibble containing the results of t-tests and log2fc compared to time 0 14,497 × 19*

Description

log2fc_t_test_data A tibble containing the results of t-tests and log2fc compared to time 0 14,497 × 19

Usage

```
log2fc_t_test_data
```

Format

'peaks_prepared_data' A tibble: 14,497 × 19

Source

<mspms processed data originally from PEAKS files found in "tests/testdata/protein-peptides-id.csv" and "tests/testdata/protein-peptides-lfq.csv">

`mspms_tidy` *mspms_tidy Convert a SummarizedExperiment object within a QFeatures object into a tidy tibble.*

Description

mspms_tidy Convert a SummarizedExperiment object within a QFeatures object into a tidy tibble.

Usage

```
mspms_tidy(processed_qf, se_name = "peptides_norm")
```

Arguments

`processed_qf`	a QFeature object containing rowData and colData.
`se_name`	the name of the SummarizedExperiment you would like to extract

Value

a tibble containing all the rowData, colData, and assay data for the specified SummarizedExperiment.

Examples

```
mspms_data <- mspms_tidy(mspms::processed_qf)
```

`mspms_tidy_data` *mspms_tidy_data A tibble containing tidy data derived from QFeatures object*

Description

mspms_tidy_data A tibble containing tidy data derived from QFeatures object

Usage

```
mspms_tidy_data
```

Format

‘mspms_tidy_data‘ A tibble:

Source

processed_qf

`peaks_prepared_data` *peaks_prepared_data A QFeatures object prepared from PEAKS data of cathepsin data/.*

Description

peaks_prepared_data A QFeatures object prepared from PEAKS data of cathepsin data/.

Usage

```
peaks_prepared_data
```

Format

‘peaks_prepared_data‘ An instance of class QFeatures containing 1 assays: [1] peptides: SummarizedExperiment with 2071 rows and 42 columns

peptides Peptide Sequence Detected ...

Source

<mspms processed data originally from PEAKS files found in "tests/testdata/protein-peptides-id.csv" and "tests/testdata/protein-peptides-lfq.csv">

peptide_library *peptide_library*

Description

This is the 228 peptide library used by the O'Donoghue lab as of 26April2024.

Usage

```
peptide_library
```

Format

'peptide_library' A data frame with 228 rows and 3 columns:

library_reference_id reference id of the detected peptide as put in upstream software

library_match_sequence the sequence match to the peptide library, methionine is replaced with norleucine,which should function the same as methionine for proteases but has the same mass as L

library_real_sequence Ls corresponding to norleucine are replaced back with n (for norleucine) ...

Source

<O'Donoghue lab as of 26April2024 >

plot_all_icelogos *plot_all_icelogos*

Description

Easily plot a iceLogo corresponding to peptides of interest across each condition of an experiment.

Usage

```
plot_all_icelogos(
  sig_cleavage_data,
  type = "percent_difference",
  pval = 0.05,
  background_universe = mspms::all_possible_8mers_from_228_library
)
```

Arguments

sig_cleavage_data	a tibble of data of interest containing a column labeled peptide, cleavage_seq, and condition
type	this is the type of iceLogo you would like to generate, can be either "percent_difference" or "fold_change".
pval	this is the pvalue threshold (<=) to consider significant when determining the significance of the sig_cleavages relative to the background at each position of the iceLogo.
background_universe	this is a list cleavages you would like to compare to as background of the iceLogo

Value

a ggplot object that shows the motif of the cleavage sequences

Examples

```
# Determining cleavages of interest
sig_cleavage_data <- mspms::log2fc_t_test_data %>%
  dplyr::filter(p.adj <= 0.05, log2fc > 3)
# Plotting a iceLogo for each condition.
plot_all_icelogos(sig_cleavage_data)
```

---

plot_cleavages_per_pos

plot_cleavages_per_pos

Description

plot the number of cleavages at each

Usage

```
plot_cleavages_per_pos(sig_cleavage_data, ncol = NULL)
```

Arguments

sig_cleavage_data	a tibble of data of interest containing a column labeled peptide, cleavage_seq, condition, and cleavage_pos.
ncol	the number of columns to plot.

Value

a ggplot2 object

Examples

```
# Defining the significant peptides
sig_cleavage_data <- log2fc_t_test_data %>%
  dplyr::filter(p.adj <= 0.05, log2fc > 3)
# Plotting
p1 <- mspms::plot_cleavages_per_pos(sig_cleavage_data)
p1
```

`plot_heatmap` *plot_heatmap*

Description

This produces a heatmaply interactive heatmap of the QFeatures object with color bars representing the condition and time for each sample in each row.

Usage

```
plot_heatmap(
  mspms_tidy_data,
  value_colname = "peptides_norm",
  scale = "column",
  plot_method = "plotly"
)
```

Arguments

`mspms_tidy_data`	tidy mspms data (prepared from QFeatures object by mspms_tidy())
`value_colname`	the name of the column containing values.
`scale`	how would you like the data scaled? default is none, but can also be "row", "column", or "none"
`plot_method`	what plot method would you like to use, can use plotly or ggplot2.

Details

Each column has a colored bar representing whether the peptide is a cleavage product or a full length member of the peptide library.

Value

a heatmaply interactive heatmap

Examples

```
plot_heatmap(mspms::mspms_tidy_data)
```

plot_icelogo *plot_icelogo*

Description

This function plots the cleavage motifs that were enriched relative to background as implemented in the iceLogo method. https://iomics.ugent.be/icelogoserver/resources/manual.pdf

Usage

```
plot_icelogo(
  cleavage_seqs,
  background_universe = mspms::all_possible_8mers_from_228_library,
  pval = 0.05,
  type = "percent_difference"
)
```

Arguments

cleavage_seqs	these are the cleavage sequences of interest
background_universe	this is a list of cleavage sequences to use as the background in building the iceLogo.
pval	this is the pvalue threshold (<=) to consider significant when determining the significance of the sig_cleavages relative to the background at each position of the iceLogo.
type	this is the type of visualization you would like to perform, accepted values are either "percent_difference" or "fold_change".

Value

a ggplot2 object

Examples

```
# Determining significant cleavages for catA
catA_sig_cleavages <- mspms::log2fc_t_test_data %>%
  dplyr::filter(p.adj <= 0.05, log2fc > 3) %>%
  dplyr::filter(condition == "CatA") %>%
  dplyr::pull(cleavage_seq) %>%
  unique()

# Plotting icelogo
plot_icelogo(catA_sig_cleavages,
  background_universe = all_possible_8mers_from_228_library
)
```

plot_nd_peptides *plot_nd_peptides*

Description

plot the percentage of samples each peptide from library was undetected in (if the percentage is > 0).

Usage

```
plot_nd_peptides(
  processed_qf,
  peptide_library_ids = mspms::peptide_library$library_id
)
```

Arguments

processed_qf	a QFeatures object containing a SummarizedExperiment named "peptides"
peptide_library_ids	a vector of all peptide library ids in the experiment.

Value

a ggplot2 object

Examples

```
plot_nd_peptides(mspms::processed_qf)
```

plot_pca *plot_pca*

Description

Easily create a PCA plot from a QFeatures object containing mspms data. Ellipses are drawn around the points at a 95 Shape and colors are user specified.

Usage

```
plot_pca(
  mspms_tidy_data,
  value_colname = "peptides_norm",
  color = "time",
  shape = "condition"
)
```

Arguments

mspms_tidy_data	tidy mspms data (prepared from QFeatures object by mspms_tidy)
value_colname	the name of the column containing values.
color	the name of the variable you would like to color by.
shape	the name of the variable that you would like to determine shape by.

Value

a ggplot2 object

Examples

```
plot_pca(mspms::mspms_tidy_data)
```

---

plot_qc_check *plot_qc_check plot the the percentage of the peptide library undetected in each sample per each sample group.*

Description

plot_qc_check plot the the percentage of the peptide library undetected in each sample per each sample group.

Usage

```
plot_qc_check(
  processed_qf,
  peptide_library = mspms::peptide_library$library_id,
  full_length_threshold = NULL,
  cleavage_product_threshold = NULL,
  ncol = 2
)
```

Arguments

processed_qf	QFeatures object containing a SummarizedExperiment named "peptides"
peptide_library	a vector of all peptide library ids in the experiment.
full_length_threshold	percent to use as threshold visualized as a vertical blue dashed line
cleavage_product_threshold	percent to use as a threshold visualized as a red dashed line
ncol	n columns.

Value

a ggplot2 object.

Examples

```
plot_qc_check(mspms::processed_qf)
```

`plot_time_course` *plot_time_course*

Description

Easily plot a time course of all peptides in a QFeatures object by peptide.

Usage

```
plot_time_course(
  mspms_tidy_data,
  value_colname = "peptides_norm",
  summarize_by_mean = FALSE
)
```

Arguments

`mspms_tidy_data`	tidy mspms data (prepared from QFeatures object by mspms_tidy())
`value_colname`	the name of the column containing values.
`summarize_by_mean`	whether to summarise by mean (TRUE- show error bars +- 1 standard deviation) or not (FALSE)

Value

a ggplot2 object

Examples

```
# Determining peptide of interest
max_log2fc_pep <- mspms::log2fc_t_test_data %>%
  dplyr::filter(p.adj <= 0.05, log2fc > 3) %>%
  dplyr::filter(log2fc == max(log2fc)) %>%
  dplyr::pull(peptide)

# Defining QFeatures filter
filtered <- mspms::mspms_tidy_data %>%
  dplyr::filter(peptide == max_log2fc_pep) %>%
  plot_time_course()
```

plot_volcano *plot_volcano*

Description

create a volcano plot to generate log2fc and adjusted p values for experimental conditions

Usage

```
plot_volcano(
  log2fc_t_test_data,
  log2fc_threshold = 3,
  padj_threshold = 0.05,
  facets = "grid",
  ncol = 1
)
```

Arguments

log2fc_t_test_data	a tibble containing the log2fc and adjusted p values
log2fc_threshold	the log2fc threshold that you want displayed on plot
padj_threshold	the padj threshold that you want displayed on plot
facets	how facets should be displayed. Accepted values are grid and wrap
ncol	ncol to include if facets = "wrap"

Value

a ggplot2 object

Examples

```
p1 <- mspms::plot_volcano(mspms::log2fc_t_test_data, log2fc_threshold = 3)
p1
```

prepare_fragpipe *prepare_fragpipe*

Description

Prepare a label free quantification file exported from Fragpipe for subsequent mspms analysis.

Usage

```
prepare_fragpipe(
  combined_peptide_filepath,
  colData_filepath,
  peptide_library = mspms::peptide_library,
  n_residues = 4
)
```

Arguments

combined_peptide_filepath	file path the combined_peptide.tsv file generated by FragPipe.
colData_filepath	file path to .csv file containing colData. Must have columns named "quantCols","group","condition",and "time".
peptide_library	peptide library used with experiment. Contains columns "library_id", "library_match_sequence", and "library_real_sequence".
n_residues	the number of amino acid residues before and after the cleavage site to generate a cleavage seq for.

Value

a QFeatures object containing a summarizedExperiment named "peptides"

Examples

```
fragpipe_combined_peptide <- system.file("extdata/fragpipe_combined_peptide.tsv", package = "mspms")
colData_filepath <- system.file("extdata/colData.csv", package = "mspms")
# Prepare the data
fragpipe_prepared_data <- mspms::prepare_fragpipe(fragpipe_combined_peptide, colData_filepath)
```

---

prepare_pd *prepare_pd Prepare a label free quantification file exported from Proteome Discoverer for subsequent mspms analysis.*

Description

prepare_pd Prepare a label free quantification file exported from Proteome Discoverer for subsequent mspms analysis.

Usage

```
prepare_pd(
  peptide_groups_filepath,
  colData_filepath,
  peptide_library = mspms::peptide_library,
  n_residues = 4
)
```

Arguments

peptide_groups_filepath	filepath to PeptideGroups.txt file exported from proteome discoverer.
colData_filepath	file path to .csv file containing colData. Must have columns named "quantCols","group","condition",and "time".
peptide_library	peptide library used with experiment. Contains columns "library_id", "library_match_sequence", and "library_real_sequence".
n_residues	the number of amino acid residues before and after the cleavage site to generate a cleavage seq for.

Value

a QFeatures object containing a summarizedExperiment named "peptides"

Examples

```
peptide_groups_filepath <- system.file(
  "extdata/proteome_discoverer_PeptideGroups.txt",
  package = "mspms"
)
colData_filepath <- system.file("extdata/colData.csv", package = "mspms")
```

---

prepare_peaks *prepare_peaks Prepare a label free quantification file exported from PEAKS for subsequent mspms analysis.*

Description

prepare_peaks Prepare a label free quantification file exported from PEAKS for subsequent mspms analysis.

Usage

```
prepare_peaks(
  lfq_filepath,
  colData_filepath,
  quality_threshold = 0.3,
  peptide_library = mspms::peptide_library,
  n_residues = 4
)
```

Arguments

lfq_filepath	this is the file path to a .csv file exported from PEAKS
colData_filepath	file path to .csv file containing colData. Must have columns named "quantCols","group","condition",and "time".
quality_threshold	only consider peptides with quality scores > than this threshold.
peptide_library	peptide library used in the experiment.
n_residues	the number of amino acid residues before and after the cleavage site to generate a cleavage seq for.

Value

a QFeatures object containing a summarizedExperiment named "peptides"

Examples

```
lfq_filepath <- system.file("extdata/peaks_protein-peptides-lfq.csv", package = "mspms")
colData_filepath <- system.file("extdata/colData.csv", package = "mspms")
# Prepare the data
peaks_prepared_data <- mspms::prepare_peaks(lfq_filepath, colData_filepath)
```

---

processed_qf *processed_qf A QFeatures object prepared from PEAKS data of Cathepsin data that has been processed (imputation/normalization)*

Description

processed_qf A QFeatures object prepared from PEAKS data of Cathepsin data that has been processed (imputation/normalization)

Usage

```
processed_qf
```

Format

‘peaks_prepared_data‘ An instance of class QFeatures containing 5 assays: [1] peptides: SummarizedExperiment with 2071 rows and 42 columns [2] peptides_log: SummarizedExperiment with 2071 rows and 42 columns [3] peptides_log_norm: SummarizedExperiment with 2071 rows and 42 columns [4] peptides_log_impute_norm: SummarizedExperiment with 2071 rows and 42 columns [5] peptides_norm: SummarizedExperiment with 2071 rows and 42 columns

peptides Peptide Sequence Detected ...

Source

<mspms processed data originally from PEAKS files found in "tests/testdata/protein-peptides-id.csv" and "tests/testdata/protein-peptides-lfq.csv">

`process_qf` *process_qf*

Description

process_qf

Usage

```
process_qf(prepared_qf)
```

Arguments

`prepared_qf`	this is a QFeatures object containing a SummarizedExperiment named "peptides"

Value

a QFeatures object containing a SummarizedExperiments named "peptides","peptides_log","peptides_log_norm", "peptides_log_impute_norm",and "peptides_norm"

Examples

```
processed_qf <- process_qf(mspms::peaks_prepared_data)
```

Index