Package: protag (via r-universe)

October 16, 2024

Title Search Tagged Peptides & Draw Highlighted Mass Spectra

Version 1.0.0

Description In a typical protein labelling procedure, proteins are chemically tagged with a functional group, usually at specific sites, then digested into peptides, which are then analyzed using matrix-assisted laser desorption ionization - time of flight mass spectrometry (MALDI-TOF MS) to generate peptide fingerprint. Relative to the control, peptides that are heavier by the mass of the labelling group are informative for sequence determination. Searching for peptides with such mass shifts, however, can be difficult. This package, designed to tackle this inconvenience, takes as input the mass list of two or multiple MALDI-TOF MS mass lists, and makes pairwise comparisons between the labeled groups vs. control, and restores centroid mass spectra with highlighted peaks of interest for easier visual examination. Particularly, peaks differentiated by the mass of the labelling group are defined as a "pair", those with equal masses as a "match", and all the other peaks as a "mismatch".For more bioanalytical background information, refer to following publications: Jingjing Deng (2015) <doi:10.1007/978-1-4939-2550-6_19>; Elizabeth Chang (2016) <doi:10.7171/jbt.16-2702-002>.

Depends R (>= 2.10)

Imports ggplot2, dplyr, RColorBrewer, grDevices

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Suggests testthat

NeedsCompilation no

Author Bo Yuan [aut, cre]

Maintainer Bo Yuan <yuanbo.faith@gmail.com>

Contents

---

myoglobin	*Simulated MALDI-TOF data of equine myoglobin, labeled (with dimethylation) and control*

Description

This example data is loosely based on the predicted MALDI-TOF mass list of tryptic peptides of equine myoglobin, created using entry P68082 (MYG_HORSE) in UniProtKB (https://www.uniprot.org/uniprot/P68082) with further modification. Labeled peptides, either singly or doubly dimethylated, present 28 or 56 Da shfit relative to the control.

Usage

    myoglobin

Format

a tibble dataset of 193 rows and 7 variables.

Author(s)

Bo Yuan <yuanbo.faith@gmail.com>

mass the mass of singly protonated peptides, in Dalton. Cysteins were alkylated with iodoacetate.

position the positions of the N- and C-termini of peptides, counting from signal amino acid methionine. The position number of N-terminus of mature protein starts at number 2.

no.MC number of mis-cleavages, which are cites of C-termini of lysines and arginines (without following proline) where trypsin fail to digest apart.

peptide.sequence the sequence of the tryptic peptides of myoglobin, composed of amino acids' single-letter abbreviations.

group four experimental groups, consisting "label1", "label2", "label3", and "control"

err error of the mass, in Dalton.

err.ppm error of the mass, in parts per million (ppm) .

intensity peak intensity, in "counts".

References

https://www.uniprot.org/uniprot/P68082

---

protag	*protag: Search Tagged Peptides & Draw Highlighted Mass Spectra*

Description

In a typical protein labelling procedure, proteins are chemically tagged with a functional group, usually at a specific site, digested into peptides, which are then analyzed using matrix-assisted laser desorption ionization - time of flight mass spectrometry (MALDI-TOF MS) to generate peptide fingerprint. Relative to the control, peptides that are heavier by the mass of the labelling group are informative for sequence determination. Searching for peptides with such mass shifts, however, can be difficult. This package, designed to tackle this inconvenience, takes as input the mass list of two or multiple MALDI-TOF MS mass lists, and makes pairwise comparisons between the labeled groups vs. control, and restores centroid mass spectra with highlighted peaks of interest for easier visual examination. Particularly, peaks differentiated by the mass of the labelling group are defined as a "pair", those with equal masses as a "match", and all the other peaks as a "mismatch". For more bioanalytical background information, refer to following publications: Jingjing Deng (2015) <doi:10.1007/978-1-4939-2550-6_19>; Elizabeth Chang (2016) <doi:10.7171/jbt.16-2702-002>.

Author(s)

Bo Yuan | yuanbo.faith@gmail.com

---

tag.search	*tag.search*

Description

This function takes a mass list dataset containing columns of "mass", "intensity" and "group" (contains the "control" observations) , and searches within specified error tolerance for "paired" peaks, "matched" peaks, and "mismatched" peaks. Mass spectra peaks with m/z diffrence being the designated variable "delta" (within error tolerance) are defined as a "pair", and peaks of the same m/z (within error tolerance) as "match"; otherwise defined as "mismatch".

Usage

```
tag.search(dataset, delta = NA, error.Da.pair = 0.5,
  error.Da.match = 0.5, error.ppm.pair = Inf, error.ppm.match = Inf,
  intens.log.transfrom = FALSE)
```

Arguments

dataset   a tidy dataset containing the mass list. At least three numeric columns are re-
                   quired: "mass", "intensity" and "group". The "mass" refers to m/z values; "in-
                   tensity" refers to peak height/area; "group" must contain the "control" observa-
                   tions.

delta   a single numeric value, or a numeric vector when multiple m/z difference is of
                   interest. The variable "delta" reflects the mass difference between the labelled
                   proteins/peptides vs. the non-labelled (the control), caused by the chemically-
                   labelling group.

error.Da.pair   error tolerance for the paired peaks, in Dalton; default at 0.5.

error.Da.match   error tolerance for the matched peaks, in Dalton; default at 0.5.

error.ppm.pair   error tolerance threshold for the paired peaks, in ppm. For paired peaks p and q,
                   the tolerance threshold is defined as 0.5 * (p+q) * error.ppm.pair / 10^6. When
                   the absolute difference between the measured vs. theoretical delta is lower than
                   the error tolerance, then the associated two peaks are considered a pair. The
                   default value of error.ppm.pair is Inf (positive infinite); that is, the error tolerance
                   by default is controled by error.Da.pair. When error.ppm.pair is otherwise set,
                   say at 100 (ppm), then the practical error tolerance value is the smallest of either
                   the Dalton control or ppm control. When the ppm control is more desirable than
                   Dalton control, consider setting error.Da.pair = Inf.

error.ppm.match
                   error tolerance for the matched peaks, in ppm. Error tolerance control for
                   matched peaks is similar to the case of paired peaks.

intens.log.transfrom
                   default to FALSE. If set to TRUE, peak intensities will be logarithmically trans-
                   formed. This is useful for displaying low-intensity peaks that would otherwise
                   be overshadowed and less visible in the mass spectra.

Value

a tidy dataset, with the original input dataset augmented with additional columns. The content in
the input dataset remain unchanged (though the display sequence may change).

Examples

```
search.result <- tag.search(myoglobin, delta = c(14, 28), error.Da.pair = .3)
search.result
tag.spectra.listplot(search.result)
```

```
tag.spectra.butterflyplot
```
tag.spectra.butterflyplot

Description

This function takes the output dataset from tag.search, and draw using ggplot2 the centroid mass spectra displayed in a mirrored or "butterfly" manner. Peaks from the same "pair" (with designated m/z difference) are highlighted in differentiating colors, distinguished away from peaks of the "match" (with the same m/z) and the "mismatch" (neither of the prior two cases).

Usage

```
tag.spectra.butterflyplot(search.output.list, show.peak.pair = TRUE,
  show.peak.match = TRUE, show.peak.mismatch = TRUE,
  show.annotation.pair = TRUE, show.annotation.match = TRUE,
  show.annotation.mismatch = TRUE, size.peak.pair = 2,
  size.peak.match = 1, size.peak.mismatch = 0.5, size.divider = 0.3,
  size.annotation.pair = NA, size.annotation.match = NA,
  size.annotation.mismatch = NA, size.groupname = NA,
  alpha.peak.pair = 0.8, alpha.peak.match = 0.5,
  alpha.peak.mismatch = 0.2, alpha.annotation.pair = 0.8,
  alpha.annotation.match = 0.5, alpha.annotation.mismatch = 0.2,
  color.pair = 1, color.match = ″black″, color.mismatch = ″black″,
  color.groupname = ″black″, color.divider = ″black″,
  angle.annotation = 90, angle.groupname = 90, gap.groupname = 0.1,
  gap.annotation = 0.05)
```

Arguments

search.output.list
> the output list from function tag.search

show.peak.pair if TRUE, show the paired peaks

show.peak.match
> if TRUE, show the matched peaks

show.peak.mismatch
> if TRUE, show the mismatched peaks

show.annotation.pair
> if TRUE, show the m/z annotations for the paired peaks

show.annotation.match
> if TRUE, show the m/z annotations for the mathced peaks

show.annotation.mismatch
> if TRUE, show the m/z annotations for the mismatched peaks

size.peak.pair adjust the peak width of the paired peaks. All size.xxx arguments take a numeric value, same functionality as line width or text size control in ggplot2

`size.peak.match`

        adjust the peak width of the matched peaks

`size.peak.mismatch`

        adjust the peak width of the mismatched peaks

`size.divider`   adjust divider width

`size.annotation.pair`

        adjust the m/z annotation text size for the paired peaks

`size.annotation.match`

        adjust the m/z annotation text size for the matched peaks

`size.annotation.mismatch`

        adjust the m/z annotation text size for the mismatched peaks

`size.groupname` adjust the text size for groupnames (e.g., "control", "experiment1", "experiment2", etc.).

`alpha.peak.pair`

        adjust the transparency of the paired peaks. All alpha.xxx arguments take a numeric value [0,1]

`alpha.peak.match`

        adjust the transparency of the matched peaks

`alpha.peak.mismatch`

        adjust the transparency of the mismatched peaks

`alpha.annotation.pair`

        adjust the transparency of the m/z annotations for the paired peaks

`alpha.annotation.match`

        adjust the transparency of the m/z annotations for the matched peaks

`alpha.annotation.mismatch`

        adjust the transparency of the m/z annotations for the mismatched peaks

`color.pair`   control the color for the paired peaks and the associated m/z annotations. Each pair will be of the same color, and different pairs of differentiating colors. In case of multiple mass shifts being of interest within a pair, e.g., delta = c(14, 28, 56), then peaks with m/z difference of either 14, 28 or 56, all belonging to the same pair, will be of the same color. Apart from the default color set, users could otherwise choose color from RColorBrewer palettes, e.g., color.pair = "Set1", or color.pair = "Blues".

        Colors for peaks (paired, matched, and mismatched) and the respectively associated annotations are designed to be of the same set of color for maximum clarity.

`color.match`   control the color for the matched peaks with the associated m/z annotations, with default in "black". Users may otherwise reset to different colors, e.g., color.match = "firebrick". As the matched peaks and mismatched peaks are usually of less research interest than paired peaks, the matched and mismatched peaks are respectively designed to be of monocolor.

`color.mismatch` control the color for the mismatched peaks with the associated m/z annotations, with default in "black".

`color.groupname`

        control the color for the groupnames, with default in "black".

color.divider   control the color of the central divider

angle.annotation

adjust the angle for the m/z annotations, taking a numeric value. This argument is useful to avoid annotation overlap, and is particularly handy when the plot is reoriented with coord_flip().

angle.groupname

adjust the angle of the groupnames.

gap.groupname   adjust the horizontal position of groupnames. A positive numeric value adjusts the distance between groupnames and the left bound of the mass spectra; negative values shifts the groupnames to the right side.

gap.annotation   adjust the distance between m/z annotations and the top of the peak.

Details

Though similar to tag.spectra.listplot, tag.spectra.butterflyplot is specifically designed for comparison of TWO mass spectra with the highest annotation clarity. That is, the "group" variable of the associated feeding dataset should contain only two unique levels, "control" and another named level. In case of existence of more than two levels in the "group" variable, all levels except "control" will be plotted overlapped with tag.spectra.butterflyplot; and it is recommended to use tag.spectra.listplot instead for multiple spectra drawing.

Value

a ggplot2 plot.

Examples

```
subset <-  myoglobin[myoglobin$group %in% c("control", "label1"), ]
search.result <- tag.search(subset, delta = c(14, 28), error.Da.pair = .3)
tag.spectra.butterflyplot(search.result)
```

---

tag.spectra.listplot   *tag.spectra.listplot*

Description

This function takes the output dataset from tag.search, and draw using ggplot2 the centroid mass spectra displayed in a listed manner. Peaks from the same "pair" (with designated m/z difference) are highlighted in differentiating colors, distinguished away from peaks of the "match" (with the same m/z) and the "mismatch" (neither of the prior two cases).

Usage

```
tag.spectra.listplot(search.output.list, show.peak.pair = TRUE,
  show.peak.match = TRUE, show.peak.mismatch = TRUE,
  show.annotation.pair = TRUE, show.annotation.match = TRUE,
  show.annotation.mismatch = FALSE, size.peak.pair = 2,
  size.peak.match = 1, size.peak.mismatch = 0.5, size.divider = 0.3,
  size.annotation.pair = NA, size.annotation.match = NA,
  size.annotation.mismatch = NA, size.groupname = NA,
  alpha.peak.pair = 0.8, alpha.peak.match = 0.5,
  alpha.peak.mismatch = 0.2, alpha.annotation.pair = 0.8,
  alpha.annotation.match = 0.5, alpha.annotation.mismatch = 0.2,
  color.pair = 1, color.match = "black", color.mismatch = "black",
  color.groupname = "black", color.divider = "black",
  angle.annotation = 90, angle.groupname = 90, gap.groupname = 0.02,
  gap.annotation = 0.15, peak.height.shrink = 0.7)
```

Arguments

`search.output.list`
　　　　　　the output list from function tag.search

`show.peak.pair`　if TRUE, show the paired peaks

`show.peak.match`
　　　　　　if TRUE, show the matched peaks

`show.peak.mismatch`
　　　　　　if TRUE, show the mismatched peaks

`show.annotation.pair`
　　　　　　if TRUE, show the m/z annotations for the paired peaks

`show.annotation.match`
　　　　　　if TRUE, show the m/z annotations for the mathced peaks

`show.annotation.mismatch`
　　　　　　if TRUE, show the m/z annotations for the mismatched peaks

`size.peak.pair`　adjust the peak width of the paired peaks. All size.xxx arguments take a numeric value, same functionality as line width or text size control in ggplot2

`size.peak.match`
　　　　　　adjust the peak width of the matched peaks

`size.peak.mismatch`
　　　　　　adjust the peak width of the mismatched peaks

`size.divider`　adjust divider width

`size.annotation.pair`
　　　　　　adjust the m/z annotation text size for the paired peaks

`size.annotation.match`
　　　　　　adjust the m/z annotation text size for the matched peaks

`size.annotation.mismatch`
　　　　　　adjust the m/z annotation text size for the mismatched peaks

`size.groupname`　adjust the text size for groupnames (e.g., "control", "experiment1", "experiment2", etc.).

alpha.peak.pair

> adjust the transparency of the paired peaks. All alpha.xxx arguments take a numeric value [0,1]

alpha.peak.match

> adjust the transparency of the matched peaks

alpha.peak.mismatch

> adjust the transparency of the mismatched peaks

alpha.annotation.pair

> adjust the transparency of the m/z annotations for the paired peaks

alpha.annotation.match

> adjust the transparency of the m/z annotations for the matched peaks

alpha.annotation.mismatch

> adjust the transparency of the m/z annotations for the mismatched peaks

color.pair   control the color for the paired peaks and the associated m/z annotations. Each pair will be of the same color, and different pairs of differentiating colors. In case of multiple mass shifts being of interest within a pair, e.g., delta = c(14, 28, 56), then peaks with m/z difference of either 14, 28 or 56, all belonging to the same pair, will be of the same color. Apart from the default color set, users could otherwise choose color from RColorBrewer palettes, e.g., color.pair = "Set1", or color.pair = "Blues".

> Colors for peaks (paired, matched, and mismatched) and the respectively associated annotations are designed to be of the same set of color for maximum clarity.

color.match   control the color for the matched peaks with the associated m/z annotations, with default in "black". Users may otherwise reset to different colors, e.g., color.match = "firebrick". As the matched peaks and mismatched peaks are usually of less research interest than paired peaks, the matched and mismatched peaks are respectively designed to be of monocolor.

color.mismatch   control the color for the mismatched peaks with the associated m/z annotations, with default in "black".

color.groupname

> control the color for the groupnames, with default in "black".

color.divider   control the color of the central divider

angle.annotation

> adjust the angle for the m/z annotations, taking a numeric value. This argument is useful to avoid annotation overlap, and is particularly handy when the plot is reoriented with coord_flip().

angle.groupname

> adjust the angle of the groupnames.

gap.groupname   adjust the horizontal position of groupnames. A positive numeric value adjusts the distance between groupnames and the left bound of the mass spectra; negative values shifts the groupnames to the right side.

gap.annotation   adjust the distance between m/z annotations and the top of the peak.

peak.height.shrink

> Taking a numeric value [0, 1], a small shrinking factor renders smaller peak height, and generates more space between peak and the central divider, leaving

more space for annotations. This argument resolves overlap among annotations with upper-floor-residing peaks, a problem unique to listplot. Therefore, this argument is not used in the butterflyplot.

Details

This function is designed for comparison of multiple mass spectra. In case of comparison of two mass spectra, it is recommended to use tag.spectra.butterflyplot for the highest annotation clarity.

Value

a ggplot2 plot.

Examples

```
search.result <- tag.search(myoglobin, delta = c(14, 28), error.Da.pair = .3)
search.result
tag.spectra.listplot(search.result)
```

Index