Package 'ypssc'

November 19, 2021

Type Package

Title Yeast-Proteome Secondary-Structure Calculator

Description An extension for 'NetSurfP-2.0' (Klausen et al. (2019) <doi:10.1002/prot.25674>)
which is specifically designed to analyze the results of bottom-up-proteomics that is
primarily analyzed with 'MaxQuant' (Cox, J., Mann, M. (2008) <doi:10.1038/nbt.1511>).
This tool is designed to process a large number of yeast peptides that produced as a
results of whole yeast cell-proteome digestion and provide a coherent picture of
secondary structure of proteins.

Version 1.0.0

License GPL (>= 3)

Encoding UTF-8

StagedInstall yes

RoxygenNote 7.1.1

Suggests rmarkdown, testthat (>= 3.0.0)

Imports spelling, dplyr, readxl, stringr, eulerr, Peptides, utils,
svDialogs, tcltk

Depends R (>= 2.10)

Config/testthat/edition 3

NeedsCompilation no

Author Sajad Tasharofi [aut, cph],
Shashank Kumbhare [aut, cre, cph],
Bent Petersen [aut],
Morteza Khaledi [aut],
Amir Shahmoradi [aut]

Maintainer Shashank Kumbhare <shashank.kumbhare@mavs.uta.edu>

Repository CRAN

Date/Publication 2021-11-19 06:50:02 UTC

R **topics documented:**

---

  findAlpha   *Alpha Helix Calculator*

Description

Form bottom-up proteomics data of proteins (peptides), this function determines the sections of
proteins (in percentage) with alpha-helix, structure.

Usage

```
findAlpha(pathFileInput = NULL, pathDirOutput = NULL, ...)
```

Arguments

pathFileInput   Path of the input csv file generated from MaxQuant.

                  MaxQuant is a quantitative proteomics software designed to analyze large mass-
                  spectrometric data. The input of MaxQuant is a raw file (.raw) from high-
                  resolution mass spectrometers. After analysis of the raw file in MaxQuant, the
                  program generates a folder named "combined".

                  In this folder there is another folder named "txt" which contains many files with
                  text format (.txt). One of the files called "peptides" which is the input of the
                  ypssc to calculate secondary structures. ypssc has been designed such a way
                  that can analyzed and extract information regarding the sample regardless of the
                  name that user chosen for the sample.

pathDirOutput   Path of the directory to which the output files will be generated.

...   (for developer use only)

Value

The output of the program is a csv file (.csv) that contains 5 columns, and the number of rows de-
pends on the number of proteins in the sample.

First column contains the ID of the identified alpha-helix proteins in the sample, second column
contains the number of identified amino acids from the corresponding protein, third column contains
number of identified amino acids with alpha-helix structure, fourth column contains the number of
amino acids that the protein originally has in the SSDYP, and fifth column contains the number of
amino acids with alpha-helix structure that the protein originally has in the SSDYP.

These columns should provide all information that the user needs to know about the protein and its structural information as well as structural information about the parts of the protein that has been identified in the sample.

In addition, it also generates 4 more '.csv' files.

1. The no. of proteins found in the sample.

2. The no. of peptides found in the sample.

3. The no. of amino acids for each protein in database.

4. It is the input file from MaxQuant that's been cleaned up for the sole purpose of calculating secondary structures.

See Also

findSecondary, findBeta, findChain

Examples

```
## Not run:
findAlpha( pathFileInput = "some/path/to/inputFile.csv",
           pathDirOutput = "some/path/to/outputDir/" )

findAlpha()

## End(Not run)
```

---

findBeta   *Beta Sheet Calculator*

Description

Form bottom-up proteomics data of proteins (peptides), this function determines the sections of proteins (in percentage) with beta-sheet, structure.

Usage

```
findBeta(pathFileInput = NULL, pathDirOutput = NULL, ...)
```

Arguments

pathFileInput Path of the input csv file generated from MaxQuant.

       MaxQuant is a quantitative proteomics software designed to analyze large mass-spectrometric data. The input of MaxQuant is a raw file (.raw) from high-resolution mass spectrometers. After analysis of the raw file in MaxQuant, the program generates a folder named "combined".

       In this folder there is another folder named "txt" which contains many files with text format (.txt). One of the files called "peptides" which is the input of the ypssc to calculate secondary structures. ypssc has been designed such a way that can analyzed and extract information regarding the sample regardless of the name that user chosen for the sample.

pathDirOutput Path of the directory to which the output files will be generated.

...   (for developer use only)

Value

The output of the program is a csv file (.csv) that contains 5 columns, and the number of rows depends on the number of proteins in the sample.

First column contains the ID of the identified alpha-helix proteins in the sample, second column contains the number of identified amino acids from the corresponding protein, third column contains number of identified amino acids with secondary structure, fourth column contains the number of amino acids that the protein originally has in the SSDYP, and fifth column contains the number of amino acids with beta-sheet that the protein originally has in the SSDYP.

These columns should provide all information that the user needs to know about the protein and its structural information as well as structural information about the parts of the protein that has been identified in the sample.

In addition, it also generates 4 more '.csv' files.

1. The no. of proteins found in the sample.

2. The no. of peptides found in the sample.

3. The no. of amino acids for each protein in database.

4. It is the input file from MaxQuant that's been cleaned up for the sole purpose of calculating secondary structures.

See Also

findSecondary, findAlpha, findChain

Examples

```
## Not run:
findBeta( pathFileInput = "some/path/to/inputFile.csv",
          pathDirOutput = "some/path/to/outputDir/" )

findBeta()

## End(Not run)
```

---

findChain   *Chain Calculator*

Description

Form bottom-up proteomics data of proteins (peptides), this function determines the sections of proteins (in percentage) with primary, structure.

Usage

```
findChain(pathFileInput = NULL, pathDirOutput = NULL, ...)
```

Arguments

pathFileInput Path of the input csv file generated from MaxQuant.

MaxQuant is a quantitative proteomics software designed to analyze large mass-spectrometric data. The input of MaxQuant is a raw file (.raw) from high-resolution mass spectrometers. After analysis of the raw file in MaxQuant, the program generates a folder named "combined".

In this folder there is another folder named "txt" which contains many files with text format (.txt). One of the files called "peptides" which is the input of the ypssc to calculate secondary structures. ypssc has been designed such a way that can analyzed and extract information regarding the sample regardless of the name that user chosen for the sample.

pathDirOutput Path of the directory to which the output files will be generated.

...   (for developer use only)

Value

The output of the program is a csv file (.csv) that contains 5 columns, and the number of rows depends on the number of proteins in the sample.

First column contains the ID of the identified alpha-helix proteins in the sample, second column contains the number of identified amino acids from the corresponding protein, third column contains number of identified amino acids with secondary structure, fourth column contains the number of amino acids that the protein originally has in the SSDYP, and fifth column contains the number

of amino acids in chain structure that the protein originally has in the SSDYP.

These columns should provide all information that the user needs to know about the protein and its structural information as well as structural information about the parts of the protein that has been identified in the sample.

In addition, it also generates 4 more '.csv' files.

1. The no. of proteins found in the sample.

2. The no. of peptides found in the sample.

3. The no. of amino acids for each protein in database.

4. It is the input file from MaxQuant that's been cleaned up for the sole purpose of calculating secondary structures.

See Also

findSecondary, findAlpha, findBeta

Examples

```
## Not run:
findChain( pathFileInput = "some/path/to/inputFile.csv",
           pathDirOutput = "some/path/to/outputDir/" )

findChain()

## End(Not run)
```

---

findSecondary   *Secondary Structure Calculator*

Description

Form bottom-up proteomics data of proteins (peptides), this function determines the sections of proteins (in percentage) with secondary structure like alpha-helix, beta sheet; also determines the parts that has primary structure.

Usage

```
findSecondary(pathFileInput = NULL, pathDirOutput = NULL, ...)
```

Arguments

pathFileInput   Path of the input csv file generated from MaxQuant.

MaxQuant is a quantitative proteomics software designed to analyze large mass-spectrometric data. The input of MaxQuant is a raw file (.raw) from high-resolution mass spectrometers. After analysis of the raw file in MaxQuant, the program generates a folder named "combined".

In this folder there is another folder named "txt" which contains many files with text format (.txt). One of the files called "peptides" which is the input of the ypssc to calculate secondary structures. ypssc has been designed such a way that can analyzed and extract information regarding the sample regardless of the name that user chosen for the sample.

pathDirOutput   Path of the directory to which the output files will be generated.

...   (for developer use only)

Value

The output of the program is a csv file (.csv) that contains 5 columns, and the number of rows depends on the number of proteins in the sample.

First column contains the ID of the identified alpha-helix proteins in the sample, second column contains the number of identified amino acids from the corresponding protein, third column contains number of identified amino acids with secondary structure, fourth column contains the number of amino acids that the protein originally has in the SSDYP, and fifth column contains the number of amino acids with secondary structure that the protein originally has in the SSDYP.

These columns should provide all information that the user needs to know about the protein and its structural information as well as structural information about the parts of the protein that has been identified in the sample.

In addition, it also generates 4 more '.csv' files.

1. The no. of proteins found in the sample.

2. The no. of peptides found in the sample.

3. The no. of amino acids for each protein in database.

4. It is the input file from MaxQuant that's been cleaned up for the sole purpose of calculating secondary structures.

See Also

findAlpha, findBeta, findChain

Examples

```
## Not run:
findSecondary( pathFileInput = "some/path/to/inputFile.csv",
               pathDirOutput = "some/path/to/outputDir/" )

findSecondary()

## End(Not run)
```

Index