Package 'ProtSynthesis'

May 7, 2022

Type Package

Title Calculate Fractional Protein Synthesis Rates

Version 0.1.0

Author c(person(given = ``Federico",
family = ``Martinez-Seidel",
role = c(``aut", ``cre"),
email = ``mseidel@mpimp-golm.mpg.de",
comment = c(ORCID = ``0000-0002-1410-2492"))

Depends R (>= 3.6.1), BiocGenerics (>= 0.1.12), Biobase (>= 2.5.5)

Imports IsoCorrectoR, reshape2, sva, ComplexHeatmap, circlize,
RColorBrewer, gridExtra, viridis, matrixStats, RandoDiStats,
fitdistrplus, raster, lawstat, arrangements, multcompView,
testthat, forcats, readr

Maintainer Federico Martinez-Seidel <mseidel@mpimp-golm.mpg.de>

Description This package allows to grab measure isotopolg intensities from proteomics files and turn them into accurate and correct labelled peptide/protein fractions for biological interpretation.
The package relies in the multiplexed nature of several analytical platforms to deliver optimal results but it can be used with in its bare bones if only labelled proteomics data is available.
Proteins may have been labelled with many different stable isotopes.

License BSD 2-Clause License.

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

RemoteType github

RemoteHost api.github.com

RemoteRepo ProtSynthesis

RemoteUsername MSeidelFed

RemoteRef HEAD

RemoteSha cd26454237d5127bff2eb906609bfdf7fb164660

GithubRepo ProtSynthesis

GithubUsername MSeidelFed

GithubRef HEAD

GithubSHA1 cd26454237d5127bff2eb906609bfdf7fb164660

NeedsCompilation no

R topics documented:

`AnnotateProteins` *A function to identify and compile protein enrichment information from a fed peptide subset*

Description

This function allows users to identify the parent proteins from peptide subsets, highlight the coverage of the peptides in the protein sequence, retrieve and visualize the noise-corrected enrichments (non-corrected LPFs) and return an output table that is necessary for the next function in the workflow, which corrects LPFs.

Usage

```
AnnotateProteins(
  PeptideVector,
  Treatment,
  FileName,
  Path2FASTA,
  Path2MQev,
  LabelFactor = as.factor(rep(c(rep("Control", 3), rep("Labelled", 3)), 2)),
  EnrichmentFileDir,
  cexSeq = 1,
  ProtPTMs = c("OX", "AC"),
  SeqCharLength = 70,
  GroupPeptides = FALSE,
  verbose = FALSE,
  CorrectLab = TRUE
)
```

Arguments

`PeptideVector`	Character vector with peptide identifiers as delivered by the EnrichmentSet.R function.
`Treatment`	Factor vector with the containing sample information in the same order as samples are outlined in the enrichment file.
`FileName`	Output file name. The output file is a PDF that contains boxplots featuring individual peptide non-corrected LPFs across samples and highlighted peptide sequences in the overall protein sequence provided in the FASTA file.
`Path2FASTA`	FASTA file used for the analysis of the dataset in MaxQuant.
`Path2MQev`	MaxQuant "evidence.txt" file.

LabelFactor	Defaults to one treatment and one control, each with a labelled counterpart and triplicated. Needs to be defined as a factor and in the same order as the Treatment vector.
EnrichmentFileDir	Parent directory were the enrichment file is contained.
ProtPTMs	Character vector with the PTM codes that are part of peptides in the provided dataset.
SeqCharLength	Defaults to 70. Defines the length of each line in the output protein sequences.
GroupPeptides	Defaults to FALSE. Allows to group all peptides from a single protein into mean protein enrichment.
verbose	Defaults to FALSE. When TRUE allows to visualize the progression through the function.
CorrectLab	Defaults to TRUE. Corrects labelling percentages using the mean residual "labelling" in non-labelled samples. This ensures that only noise "labelling" percentages with low standard deviation are successfully moved to zero.

Examples

```
...
```

`EnrichmentSet` *A function to apply thresholds and test statistically peptide enrichment in order to obtain good quality labelled peptide subsets*

Usage

```
EnrichmentSet(
  EnrichmentFileDir,
  Treatment,
  LabelFactor = as.factor(rep(c(rep("Control", 3), rep("Labelled", 3)), 2)),
  NLMAX = 0.05,
  LEnrLack = 0.02,
  SigLab = 0.05,
  SigLabControl = 0.05,
  SigLabTreatment = 0.05,
  CorrectLab = TRUE
)
```

Arguments

EnrichmentFileDir	Parent directory were the enrichment file is contained.
Treatment	Factor vector with the containing sample information in the same order as samples are outlined in the enrichment file.
LabelFactor	Defaults to one treatment and one control, each with a labelled counterpart and triplicated. Needs to be defined as a factor and in the same order as the Treatment vector.

NLMAX Defaults to 5

\itemLEnrLackDefaults to 2

\itemSigLabDefaults to 5

\itemSigLabControlDefaults to 5

\itemSigLabTreatmentDefaults to 5

\itemCorrectLabDefaults to TRUE. Corrects labelling percentages using the mean residual "labelling" in non-labelled samples. This ensures that only noise "labelling" percentages with low standard deviation are successfully moved to zero.

The statistical filters applied in this section are meant to remove falsely interpreted isotopolog abundances derived from "labelled" controls, this phenomenon can result from peptide coelution and contamination if the heavy isotopolog peaks with different peptides. This in turn results in an increased relative isotope abundance that does not come from the labelling experiment. This special scenario exemplifies the utility of having non-labelled controls in your samples. This function allows users to apply thresholds on residual labelling, noise and multiple parameters that leverage on the quality of the information that is delivered. The function returns a list of peptide subsets that fit the selected criteria and may be fed as input to the subsequent function in order to annotate their parent protein identities in the experimental dataset.

...

Dynamics LCMS Tracer kProteomics

`isotopeEnrichment` *A function to optimize peptide isotopolog intensities and transform them into IsoCorrectoR input files*

Description

This function allows users to grab the output "intensities.tsv" table from isotopeEnrichment.py and transform it into the correct input format for IsoCorrectoR while selecting and optimal number of isotopolog peaks per peptide entry. Optimization of the isotopolog number relies on estimating first how many isotopologs can be expected from the atomic composition of each peptide and their natural isotopic abundance and secondly from the labelling percentage in soluble amino acid pools and the number of labelled amino acid residues in each peptide sequence.

Usage

```
isotopeEnrichment(
  PyResultsDir,
  returnCSV = TRUE,
  verbose = FALSE,
  rmIsotopologs = 0,
  OptimizeIsotopologNr = FALSE,
  files2correct = NULL,
  AA4correction = NULL,
  AAinterprtFileDir,
  ElementalFileDir,
  ProtPTMs,
  LabelledSamplesNr
)
```

Arguments

PyResultsDir	Parent directory were the XX file is contained.
returnCSV	Deafults to TRUE. returns the needed IsoCorrectoR MeasurementFile.csv and MoleculeFile.csv to the working directory if set to TRUE.
verbose	Defaults to FALSE. Allows to monitor the progression through peptides in order to spot mistakes in the input files if any.
rmIsotopologs	Defaults to 0, as an example 3 will keep 3 isotopolog peaks per peptide and discard all others. Allows users to keep a specific number of isotopolog peaks from all input peptides if beyond this number all other peaks should be detrimental to data analysis, e.g., noisy mass spectra.
OptimizeIsotopologNr	Defaults to FALSE. Parameter to replace the rmIsotopologs parameter by a customization that calculates the optimal number of isotopolog peaks per peptide taking into account its molecular formula and the enrichment percentage in soluble amino acid precursors. To use it in TRUE mode all the information need to be supplied in the subsequent parameters.
files2correct	Defaults to NULL. Needs to be used if OptimizeIsotopologNr is set to TRUE. List of file directories, where each file contains the enrichments in soluble amino acids per treatment evaluated.
AA4correction	Defaults to NULL. Needs to be used if OptimizeIsotopologNr is set to TRUE. Character vector with the names of the amino acids to be used for the calculations. The amino acids must be present in the files from the previous parameter.
AAinterprtFileDir	File directory to the interpretation file that contains in one column amino acid names and in the second colum the single letter code for those amino acids.
ElementalFileDir	Directory to the elemental file to be used in IsoCorrectoR, which contains the natural isotopic abundance (NIA) of chemical elements.
ProtPTMs	Character vector with the PTM codes that are part of peptides in the provided dataset.
LabelledSamplesNr	Integer reflecting the total number of labelled samples in the dataset.

Examples

```
...
```

LPFcorrection *A function to turn non-corrected into corrected labelled peptide/protein fractions (Corr LPFs)*

Usage

```
LPFcorrection(
  InputNonCorrMat,
  files2correct = NULL,
  AA4correction = NULL,
  AAinterprtFileDir,
```

```
  ProtPTMs,
  CorrectMeans = FALSE,
  EnrBoundary = 1,
  GroupPeptides = T
)
```

Arguments

`InputNonCorrMat`

`files2correct`	Defaults to NULL. List of file directories, where each file contains the enrichments in soluble amino acids per treatment evaluated.
`AA4correction`	Defaults to NULL. Character vector with the names of the amino acids to be used for the calculations. The amino acids must be present in the files from the previous parameter.
`AAinterprtFileDir`	File directory to the interpretation file that contains in one column amino acid names and in the second colum the single letter code for those amino acids.
`ProtPTMs`	Character vector with the PTM codes that are part of peptides in the provided dataset.
`CorrectMeans`	Defaults to FALSE. Allows users to return corrected treatment means instead of individual replicates.
`EnrBoundary`	Defaults to one. Allows to define the boundarie of fractional enrichment accepted to be returned. 1 equals to 100 \itemGroupPeptidesDefaults to FALSE. Allows to group all peptides from a single protein into mean protein enrichment. This function allows users to input their non-corrected LPF matrix from "AnnotateProteins.R" and correct the fractional enrichment in individual peptides using the enrichment percentages in amino acid soluble pools from paired treatments. The function returns the corrected matrix, which can be directly used to calculate fractional synthesis rates. ... Dynamics LCMS Tracer kProteomics

Index