

Package ‘MSstatsConvert’

June 1, 2023

Title Import Data from Various Mass Spectrometry Signal Processing
Tools to MSstats Format

Version 1.11.0

Description

MSstatsConvert provides tools for importing reports of Mass Spectrometry data processing tools into R format suitable for statistical analysis using the MSstats and MSstatsTMT packages.

License Artistic-2.0

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.1

biocViews MassSpectrometry, Proteomics, Software, DataImport,
QualityControl

Depends R (>= 4.0)

Imports data.table, log4r, methods, checkmate, utils, stringi

Suggests tinytest, covr, knitr, rmarkdown

Collate 'clean_DIANN.R' 'clean_Philosopher.R' 'clean_Spectronaut.R'
'clean_SpectroMine.R' 'clean_Skyline.R'
'clean_ProteomeDiscoverer.R' 'clean_Progenesis.R'
'clean_OpenSWATH.R' 'clean_OpenMS.R' 'clean_MaxQuant.R'
'clean_DIAUmpire.R' 'MSstatsConvert_core_functions.R'
'utils_MSstatsConvert.R' 'utils_annotation.R'
'utils_balanced_design.R' 'utils_checks.R' 'utils_classes.R'
'utils_clean_features.R' 'utils_dt_operations.R'
'utils_filtering.R' 'utils_fractions.R' 'utils_logging.R'
'utils_shared_peptides.R'

VignetteBuilder knitr

git_url <https://git.bioconductor.org/packages/MSstatsConvert>

git_branch devel

git_last_commit 42c035d

git_last_commit_date 2023-04-25

Date/Publication 2023-06-01

Author Mateusz Staniak [aut, cre],
 Meena Choi [aut],
 Ting Huang [aut],
 Olga Vitek [aut]

Maintainer Mateusz Staniak <mtst@mstaniak.pl>

R topics documented:

.addFractions	4
.adjustIntensities	4
.aggregatePSMstoPeptideIons	5
.checkAnnotation	5
.checkDDA	6
.checkDuplicatedMeasurements	6
.checkMSstatsParams	7
.checkMultiRun	7
.checkOverlappedFeatures	8
.cleanByFeature	8
.cleanRawDIANN	9
.cleanRawDIAUmpire	9
.cleanRawMaxQuant	10
.cleanRawOpenMS	10
.cleanRawOpenSWATH	11
.cleanRawPD	11
.cleanRawPDMStats	12
.cleanRawPDTMT	13
.cleanRawPhilosopher	13
.cleanRawProgenesis	14
.cleanRawSkyline	15
.cleanRawSpectroMineTMT	15
.cleanRawSpectronaut	16
.countCommonFeatures	16
.fillValues	17
.filterByPattern	17
.filterByScore	18
.filterExact	19
.filterFewMeasurements	19
.filterManyColumns	20
.filterOverlapped	21
.findAvailable	21
.fixBasicColumns	22
.fixColumnTypes	22
.fixMissingValues	23
.getChannelColumns	23
.getCorrectFraction	24
.getDataTable	24

.getFullDesign	25
.getMissingRunsPerFeature	25
.getOverlappingFeatures	26
.handleFiltering	26
.handleFractions	27
.handleFractionsLF	27
.handleFractionsTMT	28
.handleIsotopicPeaks	28
.handleSharedPeptides	29
.handleSingleFeaturePerProtein	29
.logConverterOptions	30
.logSuccess	31
.makeBalancedDesign	31
.makeExactFilterMessage	32
.makeScoreFilterMessage	32
.mergeAnnotation	33
.MSstatsFormat	34
.nullAppender	34
.onLoad	35
.removeOverlappingFeatures	35
.removeSharedPeptides	36
.selectMSstatsColumns	36
.standardizeColnames	37
.summarizeMultipleMeasurements	37
.summarizeMultiplePSMs	38
as.data.frame.MSstatsValidated	38
as.data.table.MSstatsValidated	39
getDataType	39
getInputFile	40
MSstatsBalancedDesign	41
MSstatsClean	42
MSstatsConvert	45
MSstatsImport	45
MSstatsInputFiles-class	46
MSstatsLogsSettings	47
MSstatsMakeAnnotation	48
MSstatsPreprocess	49
MSstatsSaveSessionInfo	51

<code>.addFractions</code>	<i>Add a Fraction column to the output of MSstatsPreprocess</i>
----------------------------	---

Description

Add a Fraction column to the output of MSstatsPreprocess

Usage

```
.addFractions(input)
```

Arguments

input	output of MSstatsPreprocess
-------	-----------------------------

Value

data.table

<code>.adjustIntensities</code>	<i>Fix invalid intensities: infinite to NA, between 0 and 1 to 0</i>
---------------------------------	--

Description

Fix invalid intensities: infinite to NA, between 0 and 1 to 0

Usage

```
.adjustIntensities(input)
```

Arguments

input	data.table
-------	------------

Value

data.table

<code>.aggregatePSMstoPeptideIons</code>	<i>Aggregate multiple PSMs to a single peptide ion.</i>
--	---

Description

Aggregate multiple PSMs to a single peptide ion.

Usage

```
.aggregatePSMstoPeptideIons(input, feature_columns, summary_function = sum)
```

Arguments

<code>input</code>	data.table preprocessed by one of the <code>cleanRaw*</code> functions.
<code>feature_columns</code>	chr, names of columns that define features.
<code>summary_function</code>	function that will be used to aggregate intensities if needed.

Value

data.table

<code>.checkAnnotation</code>	<i>Check if the annotation is valid</i>
-------------------------------	---

Description

Check if the annotation is valid

Usage

```
.checkAnnotation(input, annotation)
```

Arguments

<code>input</code>	data processed by the <code>MSstatsClean</code>
<code>annotation</code>	annotation created by the <code>MSstatsMakeAnnotation</code> function

Value

TRUE invisibly if the annotation is correct, throws an error otherwise

<code>.checkDDA</code>	<i>Check validity of DDA data</i>
------------------------	-----------------------------------

Description

Check validity of DDA data

Usage

```
.checkDDA(input)
```

Arguments

input	data.table preprocessed by one of the <code>cleanRaw*</code> functions.
-------	---

Value

logical

logical, TRUE means that the input dataset comes from a DDA experiment

<code>.checkDuplicatedMeasurements</code>	<i>Check if there are duplicated measurements within run</i>
---	--

Description

Check if there are duplicated measurements within run

Usage

```
.checkDuplicatedMeasurements(input)
```

Arguments

input	output of <code>MSstatsPreprocess</code>
-------	--

Value

character vector of feature labels

<code>.checkMSstatsParams</code>	<i>Check validity of parameters to the MSstatsImport function.</i>
----------------------------------	--

Description

Check validity of parameters to the MSstatsImport function.

Usage

```
.checkMSstatsParams(  
  input,  
  annotation,  
  feature_columns,  
  remove_shared_peptides,  
  remove_single_feature_proteins,  
  feature_cleaning  
)
```

Value

none, throws an error if any of the assertions fail

<code>.checkMultiRun</code>	<i>Check if fractionation exists</i>
-----------------------------	--------------------------------------

Description

Check if fractionation exists

Usage

```
.checkMultiRun(input)
```

Arguments

input	output of MSstatsPreprocess
-------	-----------------------------

Value

list of two elements: `has_fractions` (logical) indicates if fractions was detected in the input dataset, `is_risky` (logical) indicates if there was a problem with detecting fractionation.

```
.checkOverlappedFeatures
```

Check if any features are measured in multiple fractions

Description

Check if any features are measured in multiple fractions

Usage

```
.checkOverlappedFeatures(input)
```

Arguments

input output of MSstatsPreprocess

Value

data.table

```
.cleanByFeature
```

Perform by-feature operations.

Description

Perform by-feature operations.

Usage

```
.cleanByFeature(input, feature_columns, cleaning_control)
```

Arguments

input data.table preprocessed by one of the cleanRaw* functions.
 feature_columns character vector of names of columns that define features.
 cleaning_control named list of two or three elements. See the documentation for MSstatsImport
 for details.

Value

data.table

.cleanRawDIANN	<i>Clean raw Diann files</i>
----------------	------------------------------

Description

Clean raw Diann files

Usage

```
.cleanRawDIANN(msstats_object, MBR = TRUE)
```

Arguments

`msstats_object` an object of class `MSstatsDIANNFiles`.
`MBR` True if analysis was done with match between runs

Value

data.table

.cleanRawDIAUmpire	<i>Clean raw DIAUmpire files</i>
--------------------	----------------------------------

Description

Clean raw DIAUmpire files

Usage

```
.cleanRawDIAUmpire(msstats_object, use_frag, use_pept)
```

Arguments

`msstats_object` Object that inherits from `MSstatsInputFiles` class.
`use_frag` TRUE will use the selected fragment for each peptide. 'Selected_fragments' column is required.
`use_pept` TRUE will use the selected fragment for each protein 'Selected_peptides' column is required.

Value

data.table

<code>.cleanRawMaxQuant</code>	<i>Clean raw output from MaxQuant</i>
--------------------------------	---------------------------------------

Description

Clean raw output from MaxQuant

Usage

```
.cleanRawMaxQuant(
  msstats_object,
  protein_id_col,
  remove_by_site = FALSE,
  channel_columns = "Reporterintensitycorrected"
)
```

Arguments

`msstats_object` object that inherits from `MSstatsInputFiles` class.
`protein_id_col` character, name of a column with names of proteins.
`remove_by_site` logical, if TRUE, proteins only identified by site will be removed.
`channel_columns` character, regular expression that identifies channel columns in TMT data.

Value

data.table

<code>.cleanRawOpenMS</code>	<i>Clean raw output from OpenMS</i>
------------------------------	-------------------------------------

Description

Clean raw output from OpenMS

Usage

```
.cleanRawOpenMS(msstats_object)
```

Arguments

`msstats_object` an object of class `MSstatsSpectroMineFiles`.

Value

data.table

<code>.cleanRawOpenSWATH</code>	<i>Clean raw OpenSWATH files</i>
---------------------------------	----------------------------------

Description

Clean raw OpenSWATH files

Usage

```
.cleanRawOpenSWATH(msstats_object)
```

Arguments

`msstats_object` an object of class `MSstatsSpectroMineFiles`.

Value

`data.table`

<code>.cleanRawPD</code>	<i>Clean raw Proteome Discoverer data</i>
--------------------------	---

Description

Clean raw Proteome Discoverer data

Usage

```
.cleanRawPD(  
  msstats_object,  
  quantification_column,  
  protein_id_column,  
  sequence_column,  
  remove_shared,  
  remove_protein_groups = TRUE,  
  intensity_columns_regexp = "Abundance"  
)
```

Arguments

`msstats_object` an object of class `MSstatsSpectroMineFiles`.

`quantification_column`

chr, name of a column used for quantification.

`protein_id_column`

chr, name of a column with protein IDs.

sequence_column chr, name of a column with peptide sequences.

remove_shared lgl, if TRUE, shared peptides will be removed.

remove_protein_groups if TRUE, proteins with numProteins > 1 will be removed.

intensity_columns_regexp regular expressions that defines intensity columns. Defaults to "Abundance", which means that columns that contain the word "Abundance" will be treated as corresponding to intensities for different channels.

Value

data.table

<i>.cleanRawPDMSstats</i>	<i>Clean raw PD output</i>
---------------------------	----------------------------

Description

Clean raw PD output

Usage

```
.cleanRawPDMSstats(
  msstats_object,
  quantification_column,
  protein_id_column,
  sequence_column,
  remove_shared
)
```

Arguments

msstats_object an object of class MSstatsSpectroMineFiles.

quantification_column chr, name of a column used for quantification.

protein_id_column chr, name of a column with protein IDs.

sequence_column chr, name of a column with peptide sequences.

remove_shared lgl, if TRUE, shared peptides will be removed.

Value

data.table

<code>.cleanRawPDTMT</code>	<i>Clean raw TMT data from Proteome Discoverer</i>
-----------------------------	--

Description

Clean raw TMT data from Proteome Discoverer

Usage

```
.cleanRawPDTMT(  
  msstats_object,  
  remove_shared = TRUE,  
  remove_protein_groups = TRUE,  
  protein_id_column = "ProteinAccessions",  
  intensity_columns_regexp = "Abundance"  
)
```

Arguments

`msstats_object` an object of class `MSstatsSpectroMineFiles`.
`remove_shared` `lgl`, if `TRUE`, shared peptides will be removed.
`remove_protein_groups`
if `TRUE`, proteins with `numProteins > 1` will be removed.
`protein_id_column`
chr, name of a column with protein IDs.
`intensity_columns_regexp`
regular expressions that defines intensity columns. Defaults to "Abundance", which means that columns that contain the word "Abundance" will be treated as corresponding to intensities for different channels.

Value

`data.table`

<code>.cleanRawPhilosopher</code>	<i>Clean raw Philosopher files</i>
-----------------------------------	------------------------------------

Description

Clean raw Philosopher files

Usage

```
.cleanRawPhilosopher(
  msstats_object,
  protein_id_col,
  peptide_id_col,
  channels,
  remove_shared_peptides
)
```

Arguments

`msstats_object` object of class `MSstatsPhilosopherFiles`

`protein_id_col` character name of a column that identifies proteins

`peptide_id_col` character name of a column that identifies peptides

`remove_shared_peptides` logical, if TRUE, shared peptides will be removed based on the `IsUnique` column from `Philosopher` output

`channel` character vector of channel labels

Value

data.table

`.cleanRawProgenesis` *Clean raw Progenesis output*

Description

Clean raw Progenesis output

Usage

```
.cleanRawProgenesis(msstats_object, runs, fix_colnames = TRUE)
```

Arguments

`msstats_object` an object of class `MSstatsSpectroMineFiles`.

`runs` chr, vector of Run labels.

`fix_colnames` lgl, if TRUE, one of the rows will be used as colnames.

Value

data.table

<code>.cleanRawSkyline</code>	<i>Clean raw data from Skyline</i>
-------------------------------	------------------------------------

Description

Clean raw data from Skyline

Usage

```
.cleanRawSkyline(msstats_object)
```

Arguments

`msstats_object` an object of class `MSstatsSpectroMineFiles`.

Value

`data.table`

<code>.cleanRawSpectroMineTMT</code>	<i>Clean raw SpectroMine TMT data</i>
--------------------------------------	---------------------------------------

Description

Clean raw SpectroMine TMT data

Usage

```
.cleanRawSpectroMineTMT(msstats_object)
```

Arguments

`msstats_object` an object of class `MSstatsSpectroMineFiles`.

Value

`data.table`

`.cleanRawSpectronaut` *Clean raw Spectronaut output.*

Description

Clean raw Spectronaut output.

Usage

`.cleanRawSpectronaut(msstats_object, intensity)`

Arguments

`msstats_object` an object of class `MSstatsSpectronautFiles`.
`intensity` chr, specifies which column will be used for Intensity.

Value

`data.table`

`.countCommonFeatures` *Get common values from two vectors of features*

Description

Get common values from two vectors of features

Usage

`.countCommonFeatures(features_1, features_2)`

Arguments

`features_1` vector of feature names
`features_2` vector of feature_names

Value

character vector of common values of `features_1` and `features_2`

<code>.fillValues</code>	<i>Set column to a single value</i>
--------------------------	-------------------------------------

Description

Set column to a single value

Usage

```
.fillValues(input, fill_list)
```

Arguments

<code>input</code>	data.table preprocessed by one of the <code>cleanRaw*</code> functions.
<code>fill_list</code>	named list, names correspond to column names, elements to values that will be used in the columns.

Value

data.table

<code>.filterByPattern</code>	<i>Handle filtering by pattern</i>
-------------------------------	------------------------------------

Description

Handle filtering by pattern

Usage

```
.filterByPattern(input, col_name, patterns, filter, drop)
```

Arguments

<code>input</code>	data.table preprocessed by one of the <code>.cleanRaw*</code> functions.
<code>col_name</code>	chr, name of the column with peptide sequences.
<code>filter</code>	lgl, if TRUE, peptides will be actually filtered.
<code>drop</code>	lgl, if TRUE, the column will be dropped.
<code>pattern</code>	chr, regular expression - matching peptides will be removed from the data.

Value

data.table

<code>.filterByScore</code>	<i>Filter PSMs / proteins by a given score column.</i>
-----------------------------	--

Description

Filter PSMs / proteins by a given score column.

Usage

```
.filterByScore(
  input,
  score_column,
  score_threshold,
  direction,
  behavior,
  handle_na = "keep",
  fill_value = NA,
  filter = TRUE,
  drop = TRUE
)
```

Arguments

<code>input</code>	data.table preprocessed by one of the <code>.cleanRaw*</code> functions.
<code>score_column</code>	chr, name of the column that contains scores.
<code>score_threshold</code>	num, values below or above this threshold will be removed from the data.
<code>direction</code>	chr, if "greater" only values above the threshold will be retained, if "smaller" - below the threshold.
<code>behavior</code>	chr, if "remove", values below/above the threshold will be removed, if "replace", they will be set to <code>fill_value</code> .
<code>fill_value</code>	if <code>behavior = "replace"</code> , values below/above the threshold will be replaced with <code>fill_value</code> . Defaults to NA.
<code>filter</code>	If TRUE, filtering will be performed.
<code>drop</code>	if TRUE, <code>score_column</code> will be removed.

Value

data.table

<code>.filterExact</code>	<i>Filter out specified symbols.</i>
---------------------------	--------------------------------------

Description

Filter out specified symbols.

Usage

```
.filterExact(  
  input,  
  col_name,  
  filter_symbols,  
  behavior,  
  fill_value,  
  filter,  
  drop  
)
```

Arguments

<code>input</code>	data.table preprocessed by one of the <code>.cleanRaw*</code> functions.
<code>col_name</code>	chr, name of the column that will be the base for filtering
<code>filter_symbols</code>	character vector of symbols that will be removed
<code>behavior</code>	chr, if "remove", values below/above the threshold will be removed, if "replace", they will be set to <code>fill_value</code> .
<code>fill_value</code>	if <code>behavior</code> = "replace", values below/above the threshold will be replaced with <code>fill_value</code> . Defaults to NA.
<code>filter</code>	lgl, if TRUE, decoy proteins will be removed from the data.
<code>drop</code>	lgl, if TRUE, column that contains decoy proteins will be dropped.

Value

data.table

<code>.filterFewMeasurements</code>	<i>Remove features with a small number of (non-missing) measurements across runs</i>
-------------------------------------	--

Description

Remove features with a small number of (non-missing) measurements across runs

Usage

```
.filterFewMeasurements(
  input,
  min_intensity,
  remove_few,
  feature_columns = NULL
)
```

Arguments

<code>input</code>	<code>data.table</code> pre-processed by one of the <code>.cleanRaw*</code> functions.
<code>min_intensity</code>	minimum intensity that will be considered non-missing.
<code>remove_few</code>	logical, if TRUE, features that have less than three measurements will be removed. If FALSE, only features with all missing runs will be removed.
<code>features_columns</code>	chr, vector of names of columns that define features.

Value

`data.table`

<code>.filterManyColumns</code>	<i>Filter rows that contain specified symbols in multiple columns.</i>
---------------------------------	--

Description

Filter rows that contain specified symbols in multiple columns.

Usage

```
.filterManyColumns(input, filter_columns, filter_symbols)
```

Arguments

<code>input</code>	<code>data.table</code> preprocessed by one of the <code>cleanRaw*</code> functions.
<code>filter_columns</code>	chr, names of columns in which elements will be matched and removed.
<code>filter_symbols</code>	chr, vector of strings. Rows with corresponding elements in <code>filter_columns</code> will be removed.

Value

`data.table`

<code>.filterOverlapped</code>	<i>Remove overlapped features</i>
--------------------------------	-----------------------------------

Description

Remove overlapped features

Usage

```
.filterOverlapped(input, summary_function, overlapped_features)
```

Arguments

input	data.table preprocessed by one of the <code>.cleanRaw*</code> functions and merged with annotation.
summary_function	summary function (mean, sum, max) that will be used to pick one feature from multiple overlapping features
overlapped_features	features that overlap.

Value

data.table

<code>.findAvailable</code>	<i>Select an available options from a set of possibilities</i>
-----------------------------	--

Description

Select an available options from a set of possibilities

Usage

```
.findAvailable(possibilities, option_set, fall_back = NULL)
```

Arguments

possibilities	possible legal values of a variable
option_set	set of values that includes one of the possibilities
fall_back	if there is none of the possibilities in option_set, or there are multiple hits, default to fall_back

Value

same as option_set, usually character

<code>.fixBasicColumns</code>	<i>Remove underscores from sequences and change intensity type to numeric</i>
-------------------------------	---

Description

Remove underscores from sequences and change intensity type to numeric

Usage

```
.fixBasicColumns(input)
```

Arguments

<code>input</code>	<code>data.table</code>
--------------------	-------------------------

Value

`data.table`

<code>.fixColumnTypes</code>	<i>Change classes of multiple columns</i>
------------------------------	---

Description

Change classes of multiple columns

Usage

```
.fixColumnTypes(
  input,
  numeric_columns = NULL,
  character_columns = NULL,
  factor_columns = NULL
)
```

Arguments

<code>input</code>	<code>data.table</code> preprocessed by one of the <code>cleanRaw*</code> functions.
<code>numeric_columns</code>	<code>chr</code> , vector of names of columns that will be converted to numeric.
<code>character_columns</code>	<code>chr</code> , vector of names of columns that will be converted to character.
<code>factor_columns</code>	<code>chr</code> , vector of names of columns that will be converted to factor.

Value

`data.table`

<code>.fixMissingValues</code>	<i>Change labels for missing values</i>
--------------------------------	---

Description

Change labels for missing values

Usage

```
.fixMissingValues(input, fix_missing = NULL)
```

Arguments

<code>input</code>	output of <code>MSstatsPreprocess</code>
<code>fix_missing</code>	missing values can be labeled by NA, 0 or both. If NULL, data were processed by Skyline, so missing values will be denoted by both NA and 0. If "na_to_zero", NA values will be replaced by 0. If "zero_to_na", 0 values will be replaced by NA

Value

data.table

<code>.getChannelColumns</code>	<i>Get intensity columns from wide-format data</i>
---------------------------------	--

Description

Get intensity columns from wide-format data

Usage

```
.getChannelColumns(col_names, ...)
```

Arguments

<code>col_names</code>	names of columns, where some of the columns store intensity value for different channels
<code>...</code>	varying number of strings that define channel columns.

Value

character vector of column names that correspond to channel intensities

<code>.getCorrectFraction</code>	<i>Get a name of fraction with the largest number of measurements or the largest average intensity</i>
----------------------------------	--

Description

Get a name of fraction with the largest number of measurements or the largest average intensity

Usage

```
.getCorrectFraction(input)
```

Arguments

input	output of MSstatsPreprocess
-------	-----------------------------

Value

character - label of the fraction that has most measurements or highest mean intensity for a given feature

<code>.getDataTable</code>	<i>Read file from a provided path or convert given data.frame to data.table</i>
----------------------------	---

Description

Read file from a provided path or convert given data.frame to data.table

Usage

```
.getDataTable(input, ...)
```

Arguments

input	report from a signal processing tool or a path to it
...	additional parameters for data.table::fread

Value

data.table

<code>.getFullDesign</code>	<i>Create a data.frame of each combination of values for given variables</i>
-----------------------------	--

Description

Create a data.frame of each combination of values for given variables

Usage

```
.getFullDesign(input, group_col, feature_col, measurement_col, is_tmt)
```

Arguments

<code>input</code>	output of MSstatsPreprocess
<code>group_col</code>	name of column in input. Combination of values of <code>feature_col</code> and <code>measurement_col</code> will be created within each unique value of this column
<code>is_tmt</code>	if TRUE, data will be treated as coming from TMT experiment.
<code>'feature_column'</code>	name of the column that labels features
<code>'measurement_col'</code>	name of a column with measurement labels - Runs in label-free case, Channels in TMT case.

Value

data.table

<code>.getMissingRunsPerFeature</code>	<i>Get names of missing runs</i>
--	----------------------------------

Description

Get names of missing runs

Usage

```
.getMissingRunsPerFeature(input)
```

Arguments

<code>input</code>	output of MSstatsPreprocess
--------------------	-----------------------------

Value

data.table

```
.getOverlappingFeatures
```

Get features that are overlapped among multiple runs

Description

Get features that are overlapped among multiple runs

Usage

```
.getOverlappingFeatures(input)
```

Arguments

input	data.table preprocessed by one of the .cleanRaw* functions and merged with annotation.
-------	--

Value

data.table

```
.handleFiltering
```

Handle PSM/proteins scores

Description

Handle PSM/proteins scores

Usage

```
.handleFiltering(input, score_filtering, exact_filtering, pattern_filtering)
```

Arguments

input	data.table preprocessed by one of the .cleanRaw* functions.
score_filtering	list of by-score filtering controls.
exact_filtering	list of exact filtering controls.
pattern_filtering	list of by-pattern filtering controls.

Value

data.table

<code>.handleFractions</code>	<i>Check if there are overlapping features and remove if needed</i>
-------------------------------	---

Description

Check if there are overlapping features and remove if needed

Usage

```
.handleFractions(input)
```

Arguments

input	data.table preprocessed by one of the <code>.cleanRaw*</code> functions and merged with annotation.
-------	---

Value

data.table

<code>.handleFractionsLF</code>	<i>Handle overlapping features</i>
---------------------------------	------------------------------------

Description

Handle overlapping features

Usage

```
.handleFractionsLF(input)
```

Arguments

input	output of <code>MSstatsPreprocess</code>
-------	--

Value

data.table

<code>.handleFractionsTMT</code>	<i>Remove peptide ions overlapped among multiple fractions of the same biological mixture</i>
----------------------------------	---

Description

Remove peptide ions overlapped among multiple fractions of the same biological mixture

Usage

```
.handleFractionsTMT(input)
```

Arguments

input	data.table preprocessed by one of the <code>.cleanRaw*</code> functions and merged with annotation.
-------	---

Value

data.table

<code>.handleIsotopicPeaks</code>	<i>Handle isotopic peaks</i>
-----------------------------------	------------------------------

Description

Handle isotopic peaks

Usage

```
.handleIsotopicPeaks(input, aggregate = FALSE)
```

Arguments

input	data.table preprocessed by one of the <code>cleanRaw*</code> functions.
aggregate	if TRUE, isotopic peaks will be summed.

Value

data.table

`.handleSharedPeptides` *Handle shared peptides.*

Description

Handle shared peptides.

Usage

```
.handleSharedPeptides(  
  input,  
  remove_shared = TRUE,  
  protein_column = "ProteinName",  
  peptide_column = "PeptideSequence"  
)
```

Arguments

`input` data.table pre-processed by one of the `.cleanRaw*` functions.
`remove_shared` `lgl`, if `TRUE`, shared peptides will be removed
`protein_column` `chr`, name of the column with names of proteins.
`peptide_column` `chr`, name of the column with peptide sequences.

Value

data.table

`.handleSingleFeaturePerProtein`
Remove proteins only identified by a single feature

Description

Remove proteins only identified by a single feature

Usage

```
.handleSingleFeaturePerProtein(input, remove_single_feature)
```

Arguments

`input` data.table pre-processed by one of the `.cleanRaw*` functions.
`remove_single_feature`
`lgl`, if `TRUE`, proteins with a single feature will be removed.

Value

data.table

.logConverterOptions *Log information about converter options*

Description

Log information about converter options

Usage

```
.logConverterOptions(
  feature_columns,
  remove_shared_peptides,
  remove_single_feature_proteins,
  feature_cleaning,
  is_tmt = FALSE
)
```

Arguments

feature_columns
character vector of names of columns that define spectral features.

remove_shared_peptides
logical, if TRUE shared peptides will be removed.

remove_single_feature_proteins
logical, if TRUE, proteins that only have one feature will be removed.

feature_cleaning
named list with maximum two (for MSstats converters) or three (for MSstatsTMT converter) elements. If `handle_few_measurements` is set to "remove", feature with less than three measurements will be removed (otherwise it should be equal to "keep"). `summarize_multiple_psms` is a function that will be used to aggregate multiple feature measurements in a run. It should return a scalar and accept an `na.rm` parameter. For MSstatsTMT converters, setting `remove_psms_with_any_missing` will remove features which have missing values in a run from that run.

is_tmt
If TRUE, the dataset comes from a TMT experiment

Value

TRUE invisibly if message was logged

.logSuccess	<i>Make a message about successful data cleaning/importing</i>
-------------	--

Description

Make a message about successful data cleaning/importing

Usage

```
.logSuccess(tool, event)
```

Arguments

tool	name of a signal processing tool
------	----------------------------------

Value

TRUE invisibly if logging was successful

.makeBalancedDesign	<i>Fill missing rows to create balanced design</i>
---------------------	--

Description

Fill missing rows to create balanced design

Usage

```
.makeBalancedDesign(input, fill_missing)
```

Arguments

input	output of MSstatsPreprocess
fill_missing	if TRUE, missing Intensities values will be added to data and marked as NA

Value

data.table

```
.makeExactFilterMessage
```

Make a message about filtering based on fixed values

Description

Make a message about filtering based on fixed values

Usage

```
.makeExactFilterMessage(col_name, filter_symbols, behavior, fill_value)
```

Arguments

<code>col_name</code>	chr, name of the column that will be the base for filtering
<code>filter_symbols</code>	character vector of symbols that will be removed
<code>behavior</code>	chr, if "remove", values below/above the threshold will be removed, if "replace", they will be set to <code>fill_value</code> .
<code>fill_value</code>	if <code>behavior = "replace"</code> , values below/above the threshold will be replaced with <code>fill_value</code> . Defaults to NA.

Value

character - message

```
.makeScoreFilterMessage
```

Make a message about filtering based on a score

Description

Make a message about filtering based on a score

Usage

```
.makeScoreFilterMessage(
  score_column,
  score_threshold,
  direction,
  behavior,
  fill_value
)
```


Arguments

<code>score_column</code>	chr, name of the column that contains scores.
<code>score_threshold</code>	num, values below or above this threshold will be removed from the data.
<code>direction</code>	chr, if "greater" only values above the threshold will be retained, if "smaller" - below the threshold.
<code>behavior</code>	chr, if "remove", values below/above the threshold will be removed, if "replace", they will be set to <code>fill_value</code> .
<code>fill_value</code>	if <code>behavior = "replace"</code> , values below/above the threshold will be replaced with <code>fill_value</code> . Defaults to NA.

Value

character - message

<code>.mergeAnnotation</code>	<i>Merge annotation with feature data</i>
-------------------------------	---

Description

Merge annotation with feature data

Usage

```
.mergeAnnotation(input, annotation)
```

Arguments

<code>annotation</code>	data.table with annotation
<code>data.table</code>	preprocessed by one of the <code>.cleanRaw*</code> functions.

Value

data.table

<code>.MSstatsFormat</code>	<i>Output format for further analysis by MSstats</i>
-----------------------------	--

Description

Output format for further analysis by MSstats

Usage

```
.MSstatsFormat(input)
```

Arguments

<code>input</code>	<code>data.table</code>
--------------------	-------------------------

Value

object of class `MSstatsValidated` that inherits from `data.frame`

<code>.nullAppender</code>	<i>log4r appender used not to write messages</i>
----------------------------	--

Description

A convenience function written to save time on checking if messages should be printed or logs should be written to a file.

Usage

```
.nullAppender(level, ...)
```

Arguments

<code>level</code>	log level
<code>...</code>	messages - ignored

Value

NULL invisibly

.onLoad	<i>Set default logging object when package is loaded</i>
---------	--

Description

Set default logging object when package is loaded

Usage

```
.onLoad(...)
```

Arguments

... ignored

Value

none, sets options called MSstatsLog and MSstatsMsg

.removeOverlappingFeatures	<i>Replace intensities of overlapped fractions with NA, keeping only one fraction</i>
----------------------------	---

Description

Replace intensities of overlapped fractions with NA, keeping only one fraction

Usage

```
.removeOverlappingFeatures(input)
```

Arguments

input output of MSstatsPreprocess

Value

data.table

```
.removeSharedPeptides
```

Remove peptides assigned to more than one protein.

Description

Remove peptides assigned to more than one protein.

Usage

```
.removeSharedPeptides(input, protein_column, peptide_column)
```

Arguments

`input` data.table pre-processed by one of the `.cleanRaw*` functions.
`protein_column` chr, name of the column with names of proteins.
`peptide_column` chr, name of the column with peptide sequences.

Value

data.table

```
.selectMSstatsColumns
```

Select columns for MSstats format

Description

Select columns for MSstats format

Usage

```
.selectMSstatsColumns(input)
```

Arguments

`input` data.table

Value

data.table

<code>.standardizeColnames</code>	<i>Change column names to match read.table/read.csv/read.delim conventions</i>
-----------------------------------	--

Description

Change column names to match read.table/read.csv/read.delim conventions

Usage

```
.standardizeColnames(col_names)
```

Arguments

<code>col_names</code>	chr, vector of column names
------------------------	-----------------------------

Value

character vector

<code>.summarizeMultipleMeasurements</code>	<i>Summarize multiple measurements per feature in a single run</i>
---	--

Description

Summarize multiple measurements per feature in a single run

Usage

```
.summarizeMultipleMeasurements(input, aggregator, feature_columns)
```

Arguments

<code>input</code>	data.table pre-processed by one of the .cleanRaw* functions.
<code>aggregator</code>	function that will be used to aggregate duplicated values.
<code>feature_columns</code>	chr, vector of names of columns that define features.

Value

data.table

```
.summarizeMultiplePSMs
```

Pick one PSM from a data.table of several PSMs.

Description

Pick one PSM from a data.table of several PSMs.

Usage

```
.summarizeMultiplePSMs(input, summary_function)
```

Arguments

input data.table preprocessed by one of the .cleanRaw* functions.
summary_function function that will be used to aggregate intensities if needed.

Value

character - label of a chosen PSM

```
as.data.frame.MSstatsValidated
```

Convert output of converters to data.frame

Description

Convert output of converters to data.frame

Usage

```
## S3 method for class 'MSstatsValidated'  
as.data.frame(x)
```

Arguments

x object of class MSstatsValidated

Value

data.frame

as.data.table.MSstatsValidated

Convert output of converters to data.table

Description

Convert output of converters to data.table

Usage

```
## S3 method for class 'MSstatsValidated'
as.data.table(x)
```

Arguments

x object of class MSstatsValidated

Value

data.tables

getDataType

Get type of dataset from an MSstatsInputFiles object.

Description

Get type of dataset from an MSstatsInputFiles object.

Usage

```
getDataType(msstats_object)
```

```
## S4 method for signature 'MSstatsInputFiles'
getDataType(msstats_object)
```

Arguments

msstats_object object that inherits from MSstatsInputFiles class.

Value

character - label of a data type. Currently, "MSstats" or "MSstatsTMT"
character "MSstats" or "MSstatsTMT".

Examples

```
evidence_path = system.file("tinytest/raw_data/MaxQuant/mq_ev.csv",
                             package = "MSstatsConvert")
pg_path = system.file("tinytest/raw_data/MaxQuant/mq_pg.csv",
                       package = "MSstatsConvert")
evidence = read.csv(evidence_path)
pg = read.csv(pg_path)
imported = MSstatsImport(list(evidence = evidence, protein_groups = pg),
                          "MSstats", "MaxQuant")

class(imported)
getDataTypes(imported) # "MSstats"
```

getInputFile	<i>Get one of files contained in an instance of MSstatsInputFiles class.</i>
--------------	--

Description

Get one of files contained in an instance of MSstatsInputFiles class.

Usage

```
getInputFile(msstats_object, file_type)

## S4 method for signature 'MSstatsInputFiles'
getInputFile(msstats_object, file_type = "input")
```

Arguments

msstats_object object that inherits from MSstatsInputFiles class.
file_type character name of a type file. Usually equal to "input".

Value

data.table
data.table

Examples

```
evidence_path = system.file("tinytest/raw_data/MaxQuant/mq_ev.csv",
                             package = "MSstatsConvert")
pg_path = system.file("tinytest/raw_data/MaxQuant/mq_pg.csv",
                       package = "MSstatsConvert")
evidence = read.csv(evidence_path)
pg = read.csv(pg_path)
imported = MSstatsImport(list(evidence = evidence, protein_groups = pg),
                          "MSstats", "MaxQuant")

class(imported)
head(getInputFile(imported, "evidence"))
```

MSstatsBalancedDesign *Creates balanced design by removing overlapping fractions and filling incomplete rows*

Description

Creates balanced design by removing overlapping fractions and filling incomplete rows

Usage

```
MSstatsBalancedDesign(
  input,
  feature_columns,
  fill_incomplete = TRUE,
  handle_fractions = TRUE,
  fix_missing = NULL
)
```

Arguments

input	data.table processed by the MSstatsPreprocess function
feature_columns	str, names of columns that define spectral features
fill_incomplete	if TRUE (default), Intensity values for missing runs will be added as NA
handle_fractions	if TRUE (default), overlapping fractions will be resolved
fix_missing	str, optional. Defaults to NULL, which means no action. If not NULL, must be one of the options: "zero_to_na" or "na_to_zero". If "zero_to_na", Intensity values equal exactly to 0 will be converted to NA. If "na_to_zero", missing values will be replaced by zeros.

Value

data.frame of class MSstatsValidated

Examples

```
unbalanced_data = system.file("tinytest/raw_data/unbalanced_data.csv",
  package = "MSstatsConvert")
unbalanced_data = data.table::as.data.table(read.csv(unbalanced_data))
balanced = MSstatsBalancedDesign(unbalanced_data,
  c("PeptideSequence", "PrecursorCharge",
    "FragmentIon", "ProductCharge"))
dim(balanced) # Now balanced has additional rows (with Intensity = NA)
# for runs that were not included in the unbalanced_data table
```

MSstatsClean

Clean files generated by a signal processing tools.

Description

Clean files generated by a signal processing tools.

Clean DIAUmpire files

Clean MaxQuant files

Clean OpenMS files

Clean OpenSWATH files

Clean Progenesis files

Clean ProteomeDiscoverer files

Clean Skyline files

Clean SpectroMine files

Clean Spectronaut files

Clean Philosopher files

Clean DIA-NN files

Usage

```
MSstatsClean(msstats_object, ...)
```

```
## S4 method for signature 'MSstatsDIAUmpireFiles'
MSstatsClean(msstats_object, use_frag, use_pept)
```

```
## S4 method for signature 'MSstatsMaxQuantFiles'
MSstatsClean(
  msstats_object,
  protein_id_col,
  remove_by_site = FALSE,
  channel_columns = "Reporterintensitycorrected"
)
```

```
## S4 method for signature 'MSstatsOpenMSFiles'
MSstatsClean(msstats_object)
```

```
## S4 method for signature 'MSstatsOpenSWATHFiles'
MSstatsClean(msstats_object)
```

```
## S4 method for signature 'MSstatsProgenesisFiles'
MSstatsClean(msstats_object, runs, fix_colnames = TRUE)
```

```
## S4 method for signature 'MSstatsProteomeDiscovererFiles'
```

```

MSstatsClean(
  msstats_object,
  quantification_column,
  protein_id_column,
  sequence_column,
  remove_shared,
  remove_protein_groups = TRUE,
  intensity_columns_regexp = "Abundance"
)

## S4 method for signature 'MSstatsSkylineFiles'
MSstatsClean(msstats_object)

## S4 method for signature 'MSstatsSpectroMineFiles'
MSstatsClean(msstats_object)

## S4 method for signature 'MSstatsSpectronautFiles'
MSstatsClean(msstats_object, intensity)

## S4 method for signature 'MSstatsPhilosopherFiles'
MSstatsClean(
  msstats_object,
  protein_id_col,
  peptide_id_col,
  channels,
  remove_shared_peptides
)

## S4 method for signature 'MSstatsDIANNFiles'
MSstatsClean(msstats_object, MBR = TRUE)

```

Arguments

<code>msstats_object</code>	object that inherits from <code>MSstatsInputFiles</code> class.
<code>...</code>	additional parameter to specific cleaning functions.
<code>use_frag</code>	TRUE will use the selected fragment for each peptide. 'Selected_fragments' column is required.
<code>use_pept</code>	TRUE will use the selected fragment for each protein 'Selected_peptides' column is required.
<code>protein_id_col</code>	character, name of a column with names of proteins.
<code>remove_by_site</code>	logical, if TRUE, proteins only identified by site will be removed.
<code>channel_columns</code>	character, regular expression that identifies channel columns in TMT data.
<code>runs</code>	chr, vector of Run labels.
<code>fix_colnames</code>	lgl, if TRUE, one of the rows will be used as colnames.
<code>quantification_column</code>	chr, name of a column used for quantification.

protein_id_column chr, name of a column with protein IDs.

sequence_column chr, name of a column with peptide sequences.

remove_shared lgl, if TRUE, shared peptides will be removed.

remove_protein_groups if TRUE, proteins with numProteins > 1 will be removed.

intensity_columns_regexp regular expressions that defines intensity columns. Defaults to "Abundance", which means that columns that contain the word "Abundance" will be treated as corresponding to intensities for different channels.

intensity chr, specifies which column will be used for Intensity.

peptide_id_col character name of a column that identifies peptides

remove_shared_peptides logical, if TRUE, shared peptides will be removed based on the IsUnique column from Philosopher output

MBR True if analysis was done with match between runs

Value

data.table

data.table

data.table

data.table

data.table

data.table

data.table

data.table

data.table

data.table

data.table

Examples

```
evidence_path = system.file("tinytest/raw_data/MaxQuant/mq_ev.csv",
                             package = "MSstatsConvert")
pg_path = system.file("tinytest/raw_data/MaxQuant/mq_pg.csv",
                       package = "MSstatsConvert")
evidence = read.csv(evidence_path)
pg = read.csv(pg_path)
imported = MSstatsImport(list(evidence = evidence, protein_groups = pg),
                          "MSstats", "MaxQuant")
cleaned_data = MSstatsClean(imported, protein_id_col = "Proteins")
head(cleaned_data)
```

MSstatsConvert	<i>MSstatsConvert: An R Package to Convert Data from Mass Spectrometry Signal Processing Tools to MSstats Format</i>
----------------	--

Description

MSstatsConvert helps convert data from different types of mass spectrometry experiments and signal processing tools to a format suitable for statistical analysis with the MSstats and MSstatsTMT packages.

Main functions

[MSstatsLogsSettings](#) for logs management, [MSstatsImport](#) for importing files created by signal processing tools, [MSstatsClean](#) for re-formatting imported files into a consistent format, [MSstatsImport](#) for preprocessing cleaned files, [MSstatsBalancedDesign](#) for handling fractions and creating balanced data.

MSstatsImport	<i>Import files from signal processing tools.</i>
---------------	---

Description

Import files from signal processing tools.

Usage

```
MSstatsImport(input_files, type, tool, tool_version = NULL, ...)
```

Arguments

<code>input_files</code>	list of paths to input files or <code>data.frame</code> objects. Interpretation of this parameter depends on values of parameters <code>type</code> and <code>tool</code> .
<code>type</code>	chr, "MSstats" or "MSstatsTMT".
<code>tool</code>	chr, name of a signal processing tool that generated input files.
<code>tool_version</code>	not implemented yet. In the future, this parameter will allow handling different versions of each signal processing tools.
<code>...</code>	optional additional parameters to <code>data.table::fread</code> .

Value

an object of class `MSstatsInputFiles`.

Examples

```
evidence_path = system.file("tinytest/raw_data/MaxQuant/mq_ev.csv",
                             package = "MSstatsConvert")
pg_path = system.file("tinytest/raw_data/MaxQuant/mq_pg.csv",
                       package = "MSstatsConvert")
evidence = read.csv(evidence_path)
pg = read.csv(pg_path)
imported = MSstatsImport(list(evidence = evidence, protein_groups = pg),
                          "MSstats", "MaxQuant")

class(imported)
head(getInputFile(imported, "evidence"))
```

MSstatsInputFiles-class

Class to model files that describe a single MS dataset.

Description

Class to model files that describe a single MS dataset.

MSstatsDIAUmpireFiles: class for DIAUmpire files.

MSstatsMaxQuantFiles: class for MaxQuant files.

MSstatsOpenMSFiles: class for OpenMS files.

MSstatsOpenSWATHFiles: class for OpenSWATH files.

MSstatsProgenesisFiles: class for Progenesis files.

MSstatsProteomeDiscovererFiles: class for ProteomeDiscoverer files.

MSstatsSkylineFiles: class for Skyline files.

MSstatsSkylineFiles: class for SpectroMine files.

MSstatsSpectronautFiles: class for Spectronaut files.

MSstatsPhilosopherFiles: class for Philosopher files.

MSstatsDIANNFiles: class for DIA-NN files.

Slots

files named list of files generated by a signal processing tools. In most cases, this will be a single file named input. In some cases, multiple files are used, for example MaxQuant outputs evidence and proteinGroups files.

type character: "MSstats" or "MSstatsTMT".

tool character: name of a signal processing tools that generated the output. Possible values are: DIAUmpire, MaxQuant, OpenMS, OpenSWATH, Progenesis, ProteomeDiscoverer, Skyline, SpectroMine, Spectronaut.

version description of a software version of the signal processing tool. Not implemented yet.

MSstatsLogsSettings *Set how MSstats will log information from data processing*

Description

Set how MSstats will log information from data processing

Usage

```
MSstatsLogsSettings(
  use_log_file = TRUE,
  append = FALSE,
  verbose = TRUE,
  log_file_path = NULL,
  base = "MSstats_log_",
  pkg_name = "MSstats"
)
```

Arguments

use_log_file	logical. If TRUE, information about data processing will be saved to a file.
append	logical. If TRUE, information about data processing will be added to an existing log file.
verbose	logical. If TRUE, information about data processing will be printed to the console.
log_file_path	character. Path to a file to which information about data processing will be saved. If not provided, such a file will be created automatically. If append = TRUE, has to be a valid path to a file.
base	start of the file name.
pkg_name	currently "MSstats", "MSstatsPTM" or "MSstatsTMT". Each package can use its own separate log settings.

Value

TRUE invisibly in case of successful logging setup.

Examples

```
# No logging and no messages
MSstatsLogsSettings(FALSE, FALSE, FALSE)
# Log, but do not display messages
MSstatsLogsSettings(TRUE, FALSE, FALSE)
# Log to an existing file
file.create("new_log.log")
MSstatsLogsSettings(TRUE, TRUE, log_file_path = "new_log.log")
# Do not log, but display messages
MSstatsLogsSettings(FALSE)
```

Description

Usage

Arguments

Value

Examples

[illegible]

MSstatsPreprocess	<i>Preprocess outputs from MS signal processing tools for analysis with MSstats</i>
-------------------	---

Description

Preprocess outputs from MS signal processing tools for analysis with MSstats

Usage

```
MSstatsPreprocess(
  input,
  annotation,
  feature_columns,
  remove_shared_peptides = TRUE,
  remove_single_feature_proteins = TRUE,
  feature_cleaning = list(remove_features_with_few_measurements = TRUE,
    summarize_multiple_psms = max),
  score_filtering = list(),
  exact_filtering = list(),
  pattern_filtering = list(),
  columns_to_fill = list(),
  aggregate_isotopic = FALSE,
  ...
)
```

Arguments

input	data.table processed by the MSstatsClean function.
annotation	annotation file generated by a signal processing tool.
feature_columns	character vector of names of columns that define spectral features.
remove_shared_peptides	logical, if TRUE shared peptides will be removed.
remove_single_feature_proteins	logical, if TRUE, proteins that only have one feature will be removed.
feature_cleaning	named list with maximum two (for MSstats converters) or three (for MSstatsTMT converter) elements. If <code>handle_few_measurements</code> is set to "remove", feature with less than three measurements will be removed (otherwise it should be equal to "keep"). <code>summarize_multiple_psms</code> is a function that will be used to aggregate multiple feature measurements in a run. It should return a scalar and accept an <code>na.rm</code> parameter. For MSstatsTMT converters, setting <code>remove_psms_with_any_missing</code> will remove features which have missing values in a run from that run.

`score_filtering`
a list of named lists that specify filtering options. Details are provided in the vignette.

`exact_filtering`
a list of named lists that specify filtering options. Details are provided in the vignette.

`pattern_filtering`
a list of named lists that specify filtering options. Details are provided in the vignette.

`columns_to_fill`
a named list of scalars. If provided, columns with names defined by the names of this list and values corresponding to its elements will be added to the output `data.frame`.

`aggregate_isotopic`
logical. If TRUE, isotopic peaks will be summed.

...
additional parameters to `data.table::fread`.

Value

`data.table`

Examples

```
evidence_path = system.file("tinytest/raw_data/MaxQuant/mq_ev.csv",
                             package = "MSstatsConvert")
pg_path = system.file("tinytest/raw_data/MaxQuant/mq_pg.csv",
                       package = "MSstatsConvert")
evidence = read.csv(evidence_path)
pg = read.csv(pg_path)
imported = MSstatsImport(list(evidence = evidence, protein_groups = pg),
                          "MSstats", "MaxQuant")
cleaned_data = MSstatsClean(imported, protein_id_col = "Proteins")
annot_path = system.file("tinytest/raw_data/MaxQuant/annotation.csv",
                          package = "MSstatsConvert")
mq_annot = MSstatsMakeAnnotation(cleaned_data, read.csv(annot_path),
                                 Run = "Rawfile")

# To filter M-peptides and oxidatin peptides
m_filter = list(col_name = "PeptideSequence", pattern = "M",
                filter = TRUE, drop_column = FALSE)
oxidation_filter = list(col_name = "Modifications", pattern = "Oxidation",
                        filter = TRUE, drop_column = TRUE)
msstats_format = MSstatsPreprocess(
  cleaned_data, mq_annot,
  feature_columns = c("PeptideSequence", "PrecursorCharge"),
  columns_to_fill = list(FragmentIon = NA, ProductCharge = NA),
  pattern_filtering = list(oxidation = oxidation_filter, m = m_filter)
)
# Output in the standard MSstats format
head(msstats_format)
```

MSstatsSaveSessionInfo
Save session information

Description

Save session information

Usage

```
MSstatsSaveSessionInfo(  
  path = NULL,  
  append = TRUE,  
  base = "MSstats_session_info_"  
)
```

Arguments

path	optional path to output file. If not provided, "MSstats_session_info" and current timestamp will be used as a file name
append	if TRUE and file given by the path parameter already exists, session info will be appended to the file
base	beginning of a file name

Value

TRUE invisibly after session info was saved

Examples

```
MSstatsSaveSessionInfo("session_info.txt")  
MSstatsSaveSessionInfo("session_info.txt", base = "MSstatsTMT_session_info_")
```

Index

* internal

- .MSstatsFormat, [34](#)
- .addFractions, [4](#)
- .adjustIntensities, [4](#)
- .aggregatePSMstoPeptideIons, [5](#)
- .checkAnnotation, [5](#)
- .checkDDA, [6](#)
- .checkDuplicatedMeasurements, [6](#)
- .checkMSstatsParams, [7](#)
- .checkMultiRun, [7](#)
- .checkOverlappedFeatures, [8](#)
- .cleanByFeature, [8](#)
- .cleanRawDIANN, [9](#)
- .cleanRawDIAUmpire, [9](#)
- .cleanRawMaxQuant, [10](#)
- .cleanRawOpenMS, [10](#)
- .cleanRawOpenSWATH, [11](#)
- .cleanRawPDMSstats, [12](#)
- .cleanRawPDTMT, [13](#)
- .cleanRawPhilosopher, [13](#)
- .cleanRawProgenesis, [14](#)
- .cleanRawSkyline, [15](#)
- .cleanRawSpectroMineTMT, [15](#)
- .cleanRawSpectronaut, [16](#)
- .countCommonFeatures, [16](#)
- .fillValues, [17](#)
- .filterByPattern, [17](#)
- .filterByScore, [18](#)
- .filterExact, [19](#)
- .filterFewMeasurements, [19](#)
- .filterManyColumns, [20](#)
- .filterOverlapped, [21](#)
- .findAvailable, [21](#)
- .fixBasicColumns, [22](#)
- .fixColumnTypes, [22](#)
- .fixMissingValues, [23](#)
- .getChannelColumns, [23](#)
- .getCorrectFraction, [24](#)
- .getDataTable, [24](#)
- .getFullDesign, [25](#)
- .getMissingRunsPerFeature, [25](#)
- .getOverlappingFeatures, [26](#)
- .handleFiltering, [26](#)
- .handleFractions, [27](#)
- .handleFractionsLF, [27](#)
- .handleFractionsTMT, [28](#)
- .handleIsotopicPeaks, [28](#)
- .handleSharedPeptides, [29](#)
- .handleSingleFeaturePerProtein, [29](#)
- .logConverterOptions, [30](#)
- .logSuccess, [31](#)
- .makeBalancedDesign, [31](#)
- .makeExactFilterMessage, [32](#)
- .makeScoreFilterMessage, [32](#)
- .mergeAnnotation, [33](#)
- .nullAppender, [34](#)
- .onLoad, [35](#)
- .removeOverlappingFeatures, [35](#)
- .removeSharedPeptides, [36](#)
- .selectMSstatsColumns, [36](#)
- .standardizeColnames, [37](#)
- .summarizeMultipleMeasurements, [37](#)
- .summarizeMultiplePSMs, [38](#)
- getData, [39](#)
- MSstatsInputFiles-class, [46](#)
- .MSstatsFormat, [34](#)
- .addFractions, [4](#)
- .adjustIntensities, [4](#)
- .aggregatePSMstoPeptideIons, [5](#)
- .checkAnnotation, [5](#)
- .checkDDA, [6](#)
- .checkDuplicatedMeasurements, [6](#)
- .checkMSstatsParams, [7](#)
- .checkMultiRun, [7](#)
- .checkOverlappedFeatures, [8](#)
- .cleanByFeature, [8](#)
- .cleanRawDIANN, [9](#)
- .cleanRawDIAUmpire, [9](#)

- .cleanRawMaxQuant, 10
- .cleanRawOpenMS, 10
- .cleanRawOpenSWATH, 11
- .cleanRawPD, 11
- .cleanRawPDMSstats, 12
- .cleanRawPDTMT, 13
- .cleanRawPhilosopher, 13
- .cleanRawProgenesis, 14
- .cleanRawSkyline, 15
- .cleanRawSpectroMineTMT, 15
- .cleanRawSpectronaut, 16
- .countCommonFeatures, 16
- .fillValues, 17
- .filterByPattern, 17
- .filterByScore, 18
- .filterExact, 19
- .filterFewMeasurements, 19
- .filterManyColumns, 20
- .filterOverlapped, 21
- .findAvailable, 21
- .fixBasicColumns, 22
- .fixColumnTypes, 22
- .fixMissingValues, 23
- .getChannelColumns, 23
- .getCorrectFraction, 24
- .getDataTable, 24
- .getFullDesign, 25
- .getMissingRunsPerFeature, 25
- .getOverlappingFeatures, 26
- .handleFiltering, 26
- .handleFractions, 27
- .handleFractionsLF, 27
- .handleFractionsTMT, 28
- .handleIsotopicPeaks, 28
- .handleSharedPeptides, 29
- .handleSingleFeaturePerProtein, 29
- .logConverterOptions, 30
- .logSuccess, 31
- .makeBalancedDesign, 31
- .makeExactFilterMessage, 32
- .makeScoreFilterMessage, 32
- .mergeAnnotation, 33
- .nullAppender, 34
- .onLoad, 35
- .removeOverlappingFeatures, 35
- .removeSharedPeptides, 36
- .selectMSstatsColumns, 36
- .standardizeColnames, 37
- .summarizeMultipleMeasurements, 37
- .summarizeMultiplePSMs, 38
- as.data.frame.MSstatsValidated, 38
- as.data.table.MSstatsValidated, 39
- getDataTypes, 39
- getDataTypes,MSstatsInputFiles-method
(getDataTypes), 39
- getInputFile, 40
- getInputFile,MSstatsInputFiles-method
(getInputFile), 40
- MSstatsBalancedDesign, 41, 45
- MSstatsClean, 42, 45
- MSstatsClean,MSstatsDIANNFiles-method
(MSstatsClean), 42
- MSstatsClean,MSstatsDIAUmpireFiles-method
(MSstatsClean), 42
- MSstatsClean,MSstatsMaxQuantFiles-method
(MSstatsClean), 42
- MSstatsClean,MSstatsOpenMSFiles-method
(MSstatsClean), 42
- MSstatsClean,MSstatsOpenSWATHFiles-method
(MSstatsClean), 42
- MSstatsClean,MSstatsPhilosopherFiles-method
(MSstatsClean), 42
- MSstatsClean,MSstatsProgenesisFiles-method
(MSstatsClean), 42
- MSstatsClean,MSstatsProteomeDiscovererFiles-method
(MSstatsClean), 42
- MSstatsClean,MSstatsSkylineFiles-method
(MSstatsClean), 42
- MSstatsClean,MSstatsSpectroMineFiles-method
(MSstatsClean), 42
- MSstatsClean,MSstatsSpectronautFiles-method
(MSstatsClean), 42
- MSstatsConvert, 45
- MSstatsDIANNFiles-class
(MSstatsInputFiles-class), 46
- MSstatsDIAUmpireFiles-class
(MSstatsInputFiles-class), 46
- MSstatsImport, 45, 45
- MSstatsInputFiles-class, 46
- MSstatsLogsSettings, 45, 47
- MSstatsMakeAnnotation, 48
- MSstatsMaxQuantFiles-class
(MSstatsInputFiles-class), 46

MSstatsOpenMSFiles-class
 (MSstatsInputFiles-class), [46](#)
MSstatsOpenSWATHFiles-class
 (MSstatsInputFiles-class), [46](#)
MSstatsPhilosopherFiles-class
 (MSstatsInputFiles-class), [46](#)
MSstatsPreprocess, [49](#)
MSstatsProgenesisFiles-class
 (MSstatsInputFiles-class), [46](#)
MSstatsProteomeDiscovererFiles-class
 (MSstatsInputFiles-class), [46](#)
MSstatsSaveSessionInfo, [51](#)
MSstatsSkylineFiles-class
 (MSstatsInputFiles-class), [46](#)
MSstatsSpectroMineFiles-class
 (MSstatsInputFiles-class), [46](#)
MSstatsSpectronautFiles-class
 (MSstatsInputFiles-class), [46](#)