

# Package ‘MSstatsConvert’

June 22, 2025

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

**Version** 1.19.0

**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** 7.3.2

**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\_ProteinProspector.R' 'clean\_Metamorpheus.R'  
'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'  
'converters\_DIANNtoMSstatsFormat.R'  
'converters\_DIAUmpiretoMSstatsFormat.R'  
'converters\_FragPipetoMSstatsFormat.R'  
'converters\_MaxQtoMSstatsFormat.R'  
'converters\_MetamorpheusToMSstatsFormat.R'  
'converters\_OpenMStoMSstatsFormat.R'  
'converters\_OpenSWATHtoMSstatsFormat.R'  
'converters\_PDtoMSstatsFormat.R'  
'converters\_ProgenesistoMSstatsFormat.R'  
'converters\_ProteinProspectortoMSstatsTMTFormat.R'  
'converters\_SkylinetoMSstatsFormat.R'  
'converters\_SpectronauttoMSstatsFormat.R'  
'utils\_MSstatsConvert.R' 'utils\_annotation.R'  
'utils\_balanced\_design.R' 'utils\_checks.R' 'utils\_classes.R'

'utils\_clean\_features.R' 'utils\_documentation.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** 5b61e60

**git\_last\_commit\_date** 2025-04-15

**Repository** Bioconductor 3.22

**Date/Publication** 2025-06-22

**Author** Mateusz Staniak [aut, cre],  
 Devon Kohler [aut],  
 Anthony Wu [aut],  
 Meena Choi [aut],  
 Ting Huang [aut],  
 Olga Vitek [aut]

**Maintainer** Mateusz Staniak <mtst@mstaniak.pl>

## Contents

|                                        |    |
|----------------------------------------|----|
| .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 |
| .cleanRawMetamorpheus . . . . .        | 10 |
| .cleanRawOpenMS . . . . .              | 11 |
| .cleanRawOpenSWATH . . . . .           | 11 |
| .cleanRawPD . . . . .                  | 12 |
| .cleanRawPDMSstats . . . . .           | 12 |
| .cleanRawPDTMT . . . . .               | 13 |
| .cleanRawPhilosopher . . . . .         | 14 |
| .cleanRawProgenesis . . . . .          | 15 |
| .cleanRawSkyline . . . . .             | 15 |
| .cleanRawSpectroMineTMT . . . . .      | 16 |
| .cleanRawSpectronaut . . . . .         | 16 |
| .countCommonFeatures . . . . .         | 17 |
| .fillValues . . . . .                  | 17 |
| .filterByPattern . . . . .             | 18 |
| .filterByScore . . . . .               | 18 |
| .filterExact . . . . .                 | 19 |

|                                            |    |
|--------------------------------------------|----|
| .filterFewMeasurements . . . . .           | 20 |
| .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 . . . . .                      | 30 |
| .makeBalancedDesign . . . . .              | 31 |
| .makeExactFilterMessage . . . . .          | 31 |
| .makeScoreFilterMessage . . . . .          | 32 |
| .mergeAnnotation . . . . .                 | 32 |
| .MSstatsFormat . . . . .                   | 33 |
| .nullAppender . . . . .                    | 33 |
| .onLoad . . . . .                          | 34 |
| .removeOverlappingFeatures . . . . .       | 34 |
| .removeSharedPeptides . . . . .            | 35 |
| .selectMSstatsColumns . . . . .            | 35 |
| .sharedParametersAmongConverters . . . . . | 36 |
| .standardizeColnames . . . . .             | 37 |
| .summarizeMultipleMeasurements . . . . .   | 37 |
| .summarizeMultiplePSMs . . . . .           | 38 |
| as.data.frame.MSstatsValidated . . . . .   | 38 |
| as.data.table.MSstatsValidated . . . . .   | 39 |
| DIANNtoMSstatsFormat . . . . .             | 39 |
| DIAUmpiretoMSstatsFormat . . . . .         | 41 |
| FragPipetoMSstatsFormat . . . . .          | 43 |
| getDataType . . . . .                      | 44 |
| getInputFile . . . . .                     | 45 |
| MaxQtoMSstatsFormat . . . . .              | 45 |
| MetamorpheusToMSstatsFormat . . . . .      | 47 |
| MSstatsBalancedDesign . . . . .            | 49 |
| MSstatsClean . . . . .                     | 50 |
| MSstatsConvert . . . . .                   | 53 |
| MSstatsImport . . . . .                    | 53 |
| MSstatsInputFiles-class . . . . .          | 54 |
| MSstatsLogsSettings . . . . .              | 55 |
| MSstatsMakeAnnotation . . . . .            | 56 |

|                                               |    |
|-----------------------------------------------|----|
| MSstatsPreprocess . . . . .                   | 57 |
| MSstatsSaveSessionInfo . . . . .              | 59 |
| OpenMStoMSstatsFormat . . . . .               | 59 |
| OpenSWATHtoMSstatsFormat . . . . .            | 61 |
| PDtoMSstatsFormat . . . . .                   | 62 |
| ProgenesistoMSstatsFormat . . . . .           | 64 |
| ProteinProspectortoMSstatsTMTFormat . . . . . | 65 |
| SkylinetoMSstatsFormat . . . . .              | 67 |
| SpectronautoMSstatsFormat . . . . .           | 68 |

|              |           |
|--------------|-----------|
| <b>Index</b> | <b>71</b> |
|--------------|-----------|

---

|                            |                                                                 |
|----------------------------|-----------------------------------------------------------------|
| <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

---

`.aggregatePSMstoPeptideIons`*Aggregate multiple PSMs to a single peptide ion.*

---

**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

---

`.checkAnnotation`      *Check if the annotation is valid*

---

**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

|                    |                             |
|--------------------|-----------------------------|
| <code>input</code> | 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



---

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

---

**Description**

Clean raw Diann files

**Usage**

```
.cleanRawDIANN(  
  msstats_object,  
  MBR = TRUE,  
  quantificationColumn = "FragmentQuantCorrected"  
)
```

**Arguments**

`msstats_object` an object of class `MSstatsDIANNFiles`.  
`MBR` True if analysis was done with match between runs  
`quantificationColumn` Use 'FragmentQuantCorrected' (default) column for quantified intensities. 'FragmentQuantRaw' can be used instead.

**Value**

data.table

---

|                                 |                                  |
|---------------------------------|----------------------------------|
| <code>.cleanRawDIAUmpire</code> | <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>.cleanRawMetamorpheus</code> | <i>Clean raw Metamorpheus files</i> |
|------------------------------------|-------------------------------------|

---

**Description**

Clean raw Metamorpheus files

**Usage**

```
.cleanRawMetamorpheus(msstats_object)
```

**Arguments**

`msstats_object` an object of class `MSstatsMetamorpheusFiles`.

**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`

---

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

---

**Description**

Clean raw PD output

**Usage**

```
.cleanRawPDMSstats(
  msstats_object,
  quantification_column,
  protein_id_column,
  sequence_column,
  remove_shared,
  run_column = "SpectrumFile"
)
```

**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",
  run_column = "SpectrumFile"
)
```

**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_regex`  
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`

---

`.cleanRawPhilosopher`    *Clean raw Philosopher files*

---

**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

`channels`            character vector of channel labels

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

**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

---

`.cleanRawSkyline`      *Clean raw data from Skyline*

---

### **Description**

Clean raw data from Skyline

### **Usage**

```
.cleanRawSkyline(msstats_object)
```

### **Arguments**

`msstats_object`    an object of class `MSstatsSpectroMineFiles`.

### **Value**

data.table

---

`.cleanRawSpectroMineTMT`*Clean raw SpectroMine TMT data*

---

**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

|                         |                         |
|-------------------------|-------------------------|
| <code>features_1</code> | vector of feature names |
| <code>features_2</code> | vector of feature_names |

### Value

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

---

`.fillValues`     *Set column to a single value*

---

### 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>    | <code>data.table</code> 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</code> = "replace", 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

---

*.filterFewMeasurements*

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

---

**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>            | data.table 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

---

*.filterManyColumns*

*Filter rows that contain specifed symbols in multiple columns.*

---

**Description**

Filter rows that contain specifed symbols in multiple columns.

**Usage**

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

**Arguments**

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

---

|                         |                                                                               |
|-------------------------|-------------------------------------------------------------------------------|
| <i>.fixBasicColumns</i> | <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**

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

**Value**

data.table

---

|                        |                                           |
|------------------------|-------------------------------------------|
| <i>.fixColumnTypes</i> | <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**

input                    data.table preprocessed by one of the cleanRaw\* functions.  
 numeric\_columns        chr, vector of names of columns that will be converted to numeric.  
 character\_columns      chr, vector of names of columns that will be converted to character.  
 factor\_columns        chr, vector of names of columns that will be converted to factor.

**Value**

data.table

---

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

---

**Description**

Change labels for missing values

**Usage**

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

**Arguments**

input                    output of MSstatsPreprocess  
 fix\_missing            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

---

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

---

**Description**

Get intensity columns from wide-format data

**Usage**

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

**Arguments**

col\_names                names of columns, where some of the columns store intensity value for different channels  
 ...                      varying number of strings that define channel columns.

**Value**

character vector of column names that correspond to channel intensities

---

|                            |                                                                                                        |
|----------------------------|--------------------------------------------------------------------------------------------------------|
| <i>.getCorrectFraction</i> | <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

---

|                      |                                                                                 |
|----------------------|---------------------------------------------------------------------------------|
| <i>.getDataTable</i> | <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**

|                    |                                                                                                     |
|--------------------|-----------------------------------------------------------------------------------------------------|
| <code>input</code> | 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**

|                    |                                          |
|--------------------|------------------------------------------|
| <code>input</code> | 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**

|                    |                                                                                                     |
|--------------------|-----------------------------------------------------------------------------------------------------|
| <code>input</code> | 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**

|                        |                                                                         |
|------------------------|-------------------------------------------------------------------------|
| <code>input</code>     | data.table preprocessed by one of the <code>cleanRaw*</code> functions. |
| <code>aggregate</code> | 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`     *Make a message about successful data cleaning/importing*

---

### Description

Make a message about successful data cleaning/importing

**Usage**

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

**Arguments**

|                   |                                  |
|-------------------|----------------------------------|
| <code>tool</code> | name of a signal processing tool |
|-------------------|----------------------------------|

**Value**

TRUE invisibly if logging was successful

---

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

---

**Description**

Fill missing rows to create balanced design

**Usage**

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

**Arguments**

|                           |                                                                            |
|---------------------------|----------------------------------------------------------------------------|
| <code>input</code>        | output of MSstatsPreprocess                                                |
| <code>fill_missing</code> | if TRUE, missing Intensities values will be added to data and marked as NA |

**Value**

data.table

---

|                                      |                                                             |
|--------------------------------------|-------------------------------------------------------------|
| <code>.makeExactFilterMessage</code> | <i>Make a message about filtering based on fixed values</i> |
|--------------------------------------|-------------------------------------------------------------|

---

**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</code> = "replace", values below/above the threshold will be replaced with <code>fill_value</code> . Defaults to NA. |

**Value**

character - message

---

`.mergeAnnotation`

*Merge annotation with feature data*

---

**Description**

Merge annotation with feature data

**Usage**

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



**Arguments**

|            |                                                  |
|------------|--------------------------------------------------|
| annotation | data.table with annotation                       |
| data.table | preprocessed by one of the .cleanRaw* functions. |

**Value**

data.table

---

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

---

**Description**

Output format for further analysis by MSstats

**Usage**

```
.MSstatsFormat(input)
```

**Arguments**

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

**Value**

object of class MSstatsValidated that inherits from data.frame

---

|                      |                                                  |
|----------------------|--------------------------------------------------|
| <i>.nullAppender</i> | <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**

|       |                    |
|-------|--------------------|
| level | log level          |
| ...   | messages - ignored |

**Value**

NULL invisibly

---

|                      |                                                          |
|----------------------|----------------------------------------------------------|
| <code>.onLoad</code> | <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

---

|                                         |                                                                                       |
|-----------------------------------------|---------------------------------------------------------------------------------------|
| <code>.removeOverlappingFeatures</code> | <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

---

```
.sharedParametersAmongConverters
```

*A dummy function to store shared documentation items for converters.*

---

## Description

A dummy function to store shared documentation items for converters.

## Usage

```
.sharedParametersAmongConverters()
```

## Arguments

|                                         |                                                                                                                                                                                                                   |
|-----------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>removeFewMeasurements</code>      | TRUE (default) will remove the features that have 1 or 2 measurements across runs.                                                                                                                                |
| <code>useUniquePeptide</code>           | TRUE (default) removes peptides that are assigned for more than one proteins. We assume to use unique peptide for each protein.                                                                                   |
| <code>summaryforMultipleRows</code>     | max(default) or sum - when there are multiple measurements for certain feature and certain run, use highest or sum of multiple intensities.                                                                       |
| <code>removeProtein_with1Feature</code> | TRUE will remove the proteins which have only 1 feature, which is the combination of peptide, precursor charge, fragment and charge. FALSE is default.                                                            |
| <code>removeProtein_with1Peptide</code> | TRUE will remove the proteins which have only 1 peptide and charge. FALSE is default.                                                                                                                             |
| <code>removeOxidationMpeptides</code>   | TRUE will remove the peptides including 'oxidation (M)' in modification. FALSE is default.                                                                                                                        |
| <code>removeMpeptides</code>            | TRUE will remove the peptides including 'M' sequence. FALSE is default.                                                                                                                                           |
| <code>use_log_file</code>               | logical. If TRUE, information about data processing will be saved to a file.                                                                                                                                      |
| <code>append</code>                     | logical. If TRUE, information about data processing will be added to an existing log file.                                                                                                                        |
| <code>verbose</code>                    | logical. If TRUE, information about data processing will be printed to the console.                                                                                                                               |
| <code>log_file_path</code>              | 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 <code>append = TRUE</code> , has to be a valid path to a file. |
| <code>...</code>                        | additional parameters to <code>data.table::fread</code> .                                                                                                                                                         |

---

|                                   |                                                                                |
|-----------------------------------|--------------------------------------------------------------------------------|
| <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  
`...`                              Additional arguments to be passed to or from other methods.

### 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  
 ...                    Additional arguments to be passed to or from other methods.

**Value**

data.tables

---

DIANNtoMSstatsFormat    *Import Diann files*


---

**Description**

Import Diann files

**Usage**

```
DIANNtoMSstatsFormat(
  input,
  annotation = NULL,
  global_qvalue_cutoff = 0.01,
  qvalue_cutoff = 0.01,
  pg_qvalue_cutoff = 0.01,
  useUniquePeptide = TRUE,
  removeFewMeasurements = TRUE,
  removeOxidationMpeptides = TRUE,
  removeProtein_with1Feature = TRUE,
  use_log_file = TRUE,
  append = FALSE,
  verbose = TRUE,
  log_file_path = NULL,
  MBR = TRUE,
  quantificationColumn = "FragmentQuantCorrected",
  ...
)
```

**Arguments**

|                                         |                                                                                                                                                                                                                                                                                                                            |
|-----------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>input</code>                      | name of MSstats input report from Diann, which includes feature-level data.                                                                                                                                                                                                                                                |
| <code>annotation</code>                 | name of 'annotation.txt' data which includes Condition, BioReplicate, Run.                                                                                                                                                                                                                                                 |
| <code>global_qvalue_cutoff</code>       | The qvalue cutoff for the Q.Value column, i.e. the run-specific precursor q-value. Default is 0.01.                                                                                                                                                                                                                        |
| <code>qvalue_cutoff</code>              | If MBR is false, the qvalue cutoff for the Global.Q.Value column, i.e. global precursor q-value. If MBR is true, the qvalue cutoff for the Lib.Q.Value column, i.e. the q-value for the library created after the first MBR pass. Default is 0.01.                                                                         |
| <code>pg_qvalue_cutoff</code>           | If MBR is false, the qvalue cutoff for the Global.PG.Q.Value column, i.e. the global q-value for the protein group. If MBR is true, the qvalue cutoff for the Lib.PG.Q.Value column, i.e. the protein group q-value for the library created after the first MBR pass. Run should be the same as filename. Default is 0.01. |
| <code>useUniquePeptide</code>           | should unique peptides be removed                                                                                                                                                                                                                                                                                          |
| <code>removeFewMeasurements</code>      | should proteins with few measurements be removed                                                                                                                                                                                                                                                                           |
| <code>removeOxidationMpeptides</code>   | should peptides with oxidation be removed                                                                                                                                                                                                                                                                                  |
| <code>removeProtein_with1Feature</code> | should proteins with a single feature be removed                                                                                                                                                                                                                                                                           |
| <code>use_log_file</code>               | logical. If TRUE, information about data processing will be saved to a file.                                                                                                                                                                                                                                               |
| <code>append</code>                     | logical. If TRUE, information about data processing will be added to an existing log file.                                                                                                                                                                                                                                 |
| <code>verbose</code>                    | logical. If TRUE, information about data processing will be printed to the console.                                                                                                                                                                                                                                        |
| <code>log_file_path</code>              | 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 <code>append = TRUE</code> , has to be a valid path to a file.                                                                                                          |
| <code>MBR</code>                        | True if analysis was done with match between runs                                                                                                                                                                                                                                                                          |
| <code>quantificationColumn</code>       | Use 'FragmentQuantCorrected' (default) column for quantified intensities. 'FragmentQuantRaw' can be used instead.                                                                                                                                                                                                          |
| <code>...</code>                        | additional parameters to <code>data.table::fread</code> .                                                                                                                                                                                                                                                                  |

**Value**

data.frame in the MSstats required format.

**Author(s)**

Elijah Willie



**Examples**

```

input_file_path = system.file("tinytest/raw_data/DIANN/diann_input.tsv",
                              package="MSstatsConvert")
annotation_file_path = system.file("tinytest/raw_data/DIANN/annotation.csv",
                                   package = "MSstatsConvert")
input = data.table::fread(input_file_path)
annot = data.table::fread(annotation_file_path)
output = DIAANNtoMSstatsFormat(input, annotation = annot, MBR = FALSE,
                               use_log_file = FALSE)
head(output)

```

---

DIAUmpiretoMSstatsFormat

*Import DIA-Umpire files*


---

**Description**

Import DIA-Umpire files

**Usage**

```

DIAUmpiretoMSstatsFormat(
  raw.frag,
  raw.pep,
  raw.pro,
  annotation,
  useSelectedFrag = TRUE,
  useSelectedPep = TRUE,
  removeFewMeasurements = TRUE,
  removeProtein_with1Feature = FALSE,
  summaryforMultipleRows = max,
  use_log_file = TRUE,
  append = FALSE,
  verbose = TRUE,
  log_file_path = NULL,
  ...
)

```

**Arguments**

|                 |                                                                                                |
|-----------------|------------------------------------------------------------------------------------------------|
| raw.frag        | name of FragSummary_date.xls data, which includes feature-level data.                          |
| raw.pep         | name of PeptideSummary_date.xls data, which includes selected fragments information.           |
| raw.pro         | name of ProteinSummary_date.xls data, which includes selected peptides information.            |
| annotation      | name of annotation data which includes Condition, BioReplicate, Run information.               |
| useSelectedFrag | TRUE will use the selected fragment for each peptide. 'Selected_fragments' column is required. |

|                            |                                                                                                                                                                                                     |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| useSelectedPep             | TRUE will use the selected peptide for each protein. 'Selected_peptides' column is required.                                                                                                        |
| removeFewMeasurements      | TRUE (default) will remove the features that have 1 or 2 measurements across runs.                                                                                                                  |
| removeProtein_with1Feature | TRUE will remove the proteins which have only 1 feature, which is the combination of peptide, precursor charge, fragment and charge. FALSE is default.                                              |
| summaryforMultipleRows     | max(default) or sum - when there are multiple measurements for certain feature and certain run, use highest or sum of multiple intensities.                                                         |
| 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. |
| ...                        | additional parameters to <code>data.table::fread</code> .                                                                                                                                           |

**Value**

data.frame in the MSstats required format.

**Author(s)**

Meena Choi, Olga Vitek

**Examples**

```

diau_frag = system.file("tinytest/raw_data/DIAUmpire/dia_frag.csv",
                        package = "MSstatsConvert")
diau_pept = system.file("tinytest/raw_data/DIAUmpire/dia_pept.csv",
                        package = "MSstatsConvert")
diau_prot = system.file("tinytest/raw_data/DIAUmpire/dia_prot.csv",
                        package = "MSstatsConvert")
annot = system.file("tinytest/raw_data/DIAUmpire/annot_diau.csv",
                   package = "MSstatsConvert")
diau_frag = data.table::fread(diau_frag)
diau_pept = data.table::fread(diau_pept)
diau_prot = data.table::fread(diau_prot)
annot = data.table::fread(annot)
diau_frag = diau_frag[, lapply(.SD, function(x) if (is.integer(x)) as.numeric(x) else x)]
# In case numeric columns are not interpreted correctly

diau_imported = DIAUmpiretoMSstatsFormat(diau_frag, diau_pept, diau_prot,
                                         annot, use_log_file = FALSE)

head(diau_imported)

```

---

FragPipeToMSstatsFormat

*Import FragPipe files*


---

## Description

Import FragPipe files

## Usage

```
FragPipeToMSstatsFormat(
  input,
  useUniquePeptide = TRUE,
  removeFewMeasurements = TRUE,
  removeProtein_with1Feature = FALSE,
  summaryforMultipleRows = max,
  use_log_file = TRUE,
  append = FALSE,
  verbose = TRUE,
  log_file_path = NULL,
  ...
)
```

## Arguments

|                            |                                                                                                                                                                                                     |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| input                      | name of FragPipe msstats.csv export. ProteinName, PeptideSequence, PrecursorCharge, FragmentIon, ProductCharge, IsotopeLabelType, Condition, BioReplicate, Run, Intensity are required.             |
| useUniquePeptide           | TRUE (default) removes peptides that are assigned for more than one proteins. We assume to use unique peptide for each protein.                                                                     |
| removeFewMeasurements      | TRUE (default) will remove the features that have 1 or 2 measurements across runs.                                                                                                                  |
| removeProtein_with1Feature | TRUE will remove the proteins which have only 1 feature, which is the combination of peptide, precursor charge, fragment and charge. FALSE is default.                                              |
| summaryforMultipleRows     | max(default) or sum - when there are multiple measurements for certain feature and certain run, use highest or sum of multiple intensities.                                                         |
| 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. |
| ...                        | additional parameters to <code>data.table::fread</code> .                                                                                                                                           |

**Value**

data.frame in the MSstats required format.

**Author(s)**

Devon Kohler

**Examples**

```
fragpipe_raw = system.file("tinytest/raw_data/FragPipe/fragpipe_input.csv",
                           package = "MSstatsConvert")
fragpipe_raw = data.table::fread(fragpipe_raw)
fragpipe_imported = FragPipeToMSstatsFormat(fragpipe_raw, use_log_file = FALSE)
head(fragpipe_imported)
```

---

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)
getDataType(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")

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

### Arguments

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

### Value

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")

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

---

|                     |                              |
|---------------------|------------------------------|
| MaxQtoMSstatsFormat | <i>Import MaxQuant files</i> |
|---------------------|------------------------------|

---

### Description

Import MaxQuant files

**Usage**

```

MaxQtoMSstatsFormat(
  evidence,
  annotation,
  proteinGroups,
  proteinID = "Proteins",
  useUniquePeptide = TRUE,
  summaryforMultipleRows = max,
  removeFewMeasurements = TRUE,
  removeMpeptides = FALSE,
  removeOxidationMpeptides = FALSE,
  removeProtein_with1Peptide = FALSE,
  use_log_file = TRUE,
  append = FALSE,
  verbose = TRUE,
  log_file_path = NULL,
  ...
)

```

**Arguments**

|                            |                                                                                                                                             |
|----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| evidence                   | name of 'evidence.txt' data, which includes feature-level data.                                                                             |
| annotation                 | name of 'annotation.txt' data which includes Raw.file, Condition, BioReplicate, Run, IsotopeLabelType information.                          |
| proteinGroups              | name of 'proteinGroups.txt' data. It needs to matching protein group ID. If proteinGroups=NULL, use 'Proteins' column in 'evidence.txt'.    |
| proteinID                  | 'Proteins'(default) or 'Leading.razor.protein' for Protein ID.                                                                              |
| useUniquePeptide           | TRUE (default) removes peptides that are assigned for more than one proteins. We assume to use unique peptide for each protein.             |
| summaryforMultipleRows     | max(default) or sum - when there are multiple measurements for certain feature and certain run, use highest or sum of multiple intensities. |
| removeFewMeasurements      | TRUE (default) will remove the features that have 1 or 2 measurements across runs.                                                          |
| removeMpeptides            | TRUE will remove the peptides including 'M' sequence. FALSE is default.                                                                     |
| removeOxidationMpeptides   | TRUE will remove the peptides including 'oxidation (M)' in modification. FALSE is default.                                                  |
| removeProtein_with1Peptide | TRUE will remove the proteins which have only 1 peptide and charge. FALSE is default.                                                       |
| 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.

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

**Value**

`data.frame` in the MSstats required format.

**Note**

Warning: MSstats does not support for metabolic labeling or iTRAQ experiments.

**Author(s)**

Meena Choi, Olga Vitek.

**Examples**

```
mq_ev = data.table::fread(system.file("tinytest/raw_data/MaxQuant/mq_ev.csv",
                                     package = "MSstatsConvert"))
mq_pg = data.table::fread(system.file("tinytest/raw_data/MaxQuant/mq_pg.csv",
                                     package = "MSstatsConvert"))
annot = data.table::fread(system.file("tinytest/raw_data/MaxQuant/annotation.csv",
                                     package = "MSstatsConvert"))
maxq_imported = MaxQtoMSstatsFormat(mq_ev, annot, mq_pg, use_log_file = FALSE)
head(maxq_imported)
```

---

MetamorpheusToMSstatsFormat

*Import Metamorpheus files*

---

**Description**

Import Metamorpheus files

**Usage**

```
MetamorpheusToMSstatsFormat(
  input,
  annotation = NULL,
  useUniquePeptide = TRUE,
  removeFewMeasurements = TRUE,
  removeProtein_with1Feature = FALSE,
  summaryforMultipleRows = max,
  use_log_file = TRUE,
  append = FALSE,
  verbose = TRUE,
  log_file_path = NULL,
  ...
)
```

**Arguments**

|                                         |                                                                                                                                                                                                                   |
|-----------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>input</code>                      | name of Metamorpheus output file, which is tabular format. Use the AllQuantifiedPeaks.tsv file from the Metamorpheus output.                                                                                      |
| <code>annotation</code>                 | name of 'annotation.txt' data which includes Condition, BioReplicate.                                                                                                                                             |
| <code>useUniquePeptide</code>           | TRUE (default) removes peptides that are assigned for more than one proteins. We assume to use unique peptide for each protein.                                                                                   |
| <code>removeFewMeasurements</code>      | TRUE (default) will remove the features that have 1 or 2 measurements across runs.                                                                                                                                |
| <code>removeProtein_with1Feature</code> | TRUE will remove the proteins which have only 1 feature, which is the combination of peptide, precursor charge, fragment and charge. FALSE is default.                                                            |
| <code>summaryforMultipleRows</code>     | max(default) or sum - when there are multiple measurements for certain feature and certain run, use highest or sum of multiple intensities.                                                                       |
| <code>use_log_file</code>               | logical. If TRUE, information about data processing will be saved to a file.                                                                                                                                      |
| <code>append</code>                     | logical. If TRUE, information about data processing will be added to an existing log file.                                                                                                                        |
| <code>verbose</code>                    | logical. If TRUE, information about data processing will be printed to the console.                                                                                                                               |
| <code>log_file_path</code>              | 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 <code>append = TRUE</code> , has to be a valid path to a file. |
| <code>...</code>                        | additional parameters to <code>data.table::fread</code> .                                                                                                                                                         |

**Value**

data.frame in the MSstats required format.

**Author(s)**

Anthony Wu

**Examples**

```
input = system.file("tinytest/raw_data/Metamorpheus/AllQuantifiedPeaks.tsv",
                    package = "MSstatsConvert")
input = data.table::fread(input)
annot = system.file("tinytest/raw_data/Metamorpheus/Annotation.tsv",
                    package = "MSstatsConvert")
annot = data.table::fread(annot)
metamorpheus_imported = MSstatsConvert::MetamorpheusToMSstatsFormat(input, annotation = annot)
head(metamorpheus_imported)
```



---

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,
  remove_few = TRUE
)
```

## Arguments

|                  |                                                                                                                                                                                                                                                                                                                                                                                                                       |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| input            | data.table processed by the MSstatsPreprocess function                                                                                                                                                                                                                                                                                                                                                                |
| feature_columns  | str, names of columns that define spectral features                                                                                                                                                                                                                                                                                                                                                                   |
| fill_incomplete  | if TRUE (default), ensures that rows with missing data for specific features are added as NA. For example, if the y10 ion of peptideA is measured in the "disease" samples but entirely missing for the "healthy" samples, rows with NA values will be created for the y10 ion of peptideA in the "healthy" group. This process increases the number of rows to account for all possible feature-sample combinations. |
| 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.                                                                                                                                               |
| remove_few       | lgl, if TRUE, features with one or two measurements across runs will be removed.                                                                                                                                                                                                                                                                                                                                      |

## 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

Clean Metamorpheus files

Clean Protein Prospector 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,
  quantificationColumn = "FragmentQuantCorrected"
)

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

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

```

### 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.   |



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, [MSstatsPreprocess](#) 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

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

**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.

MSstatsFragPipeFiles: class for FragPipe files.

MSstatsMetamorpheusFiles: class for Metamorpheus files.

MSstatsProteinProspectorFiles: class for ProteinProspector 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

|                      |                                                                                                                                                                                                     |
|----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>use_log_file</b>  | logical. If TRUE, information about data processing will be saved to a file.                                                                                                                        |
| <b>append</b>        | logical. If TRUE, information about data processing will be added to an existing log file.                                                                                                          |
| <b>verbose</b>       | logical. If TRUE, information about data processing will be printed to the console.                                                                                                                 |
| <b>log_file_path</b> | 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. |
| <b>base</b>          | start of the file name.                                                                                                                                                                             |
| <b>pkg_name</b>      | 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")
```

---

|                       |                   |
|-----------------------|-------------------|
| MSstatsMakeAnnotation | Create annotation |
|-----------------------|-------------------|

---

### Create annotation

```
MSstatsMakeAnnotation(input, annotation, ...)
```

|            |                                                                |
|------------|----------------------------------------------------------------|
| input      | data.table preprocessed by the MSstatsClean function           |
| annotation | data.table                                                     |
| ...        | key-value pairs, where keys are names of columns of annotation |

## data.table

[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

|                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|---------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>input</code>                          | data.table processed by the MSstatsClean function.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <code>annotation</code>                     | annotation file generated by a signal processing tool.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| <code>feature_columns</code>                | character vector of names of columns that define spectral features.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <code>remove_shared_peptides</code>         | logical, if TRUE shared peptides will be removed.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| <code>remove_single_feature_proteins</code> | logical, if TRUE, proteins that only have one feature will be removed.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| <code>feature_cleaning</code>               | 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. |
| <code>score_filtering</code>                | 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_")
```

---

`OpenMStoMSstatsFormat` *Import OpenMS files*

---

**Description**

Import OpenMS files

**Usage**

```
OpenMStoMSstatsFormat(  
  input,  
  annotation = NULL,  
  useUniquePeptide = TRUE,  
  removeFewMeasurements = TRUE,  
  removeProtein_with1Feature = FALSE,
```

```

summaryforMultipleRows = max,
use_log_file = TRUE,
append = FALSE,
verbose = TRUE,
log_file_path = NULL,
...
)

```

## Arguments

|                                         |                                                                                                                                                                                                                   |
|-----------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>input</code>                      | name of MSstats input report from OpenMS, which includes feature(peptide ion)-level data.                                                                                                                         |
| <code>annotation</code>                 | name of 'annotation.txt' data which includes Condition, BioReplicate, Run. Run should be the same as filename.                                                                                                    |
| <code>useUniquePeptide</code>           | TRUE (default) removes peptides that are assigned for more than one proteins. We assume to use unique peptide for each protein.                                                                                   |
| <code>removeFewMeasurements</code>      | TRUE (default) will remove the features that have 1 or 2 measurements across runs.                                                                                                                                |
| <code>removeProtein_with1Feature</code> | TRUE will remove the proteins which have only 1 feature, which is the combination of peptide, precursor charge, fragment and charge. FALSE is default.                                                            |
| <code>summaryforMultipleRows</code>     | max(default) or sum - when there are multiple measurements for certain feature and certain run, use highest or sum of multiple intensities.                                                                       |
| <code>use_log_file</code>               | logical. If TRUE, information about data processing will be saved to a file.                                                                                                                                      |
| <code>append</code>                     | logical. If TRUE, information about data processing will be added to an existing log file.                                                                                                                        |
| <code>verbose</code>                    | logical. If TRUE, information about data processing will be printed to the console.                                                                                                                               |
| <code>log_file_path</code>              | 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 <code>append = TRUE</code> , has to be a valid path to a file. |
| <code>...</code>                        | additional parameters to <code>data.table::fread</code> .                                                                                                                                                         |

## Value

data.frame in the MSstats required format.

## Author(s)

Meena Choi, Olga Vitek.

## Examples

```

openms_raw = data.table::fread(system.file("tinytest/raw_data/OpenMS/openms_input.csv",
                                           package = "MSstatsConvert"))
openms_imported = OpenMStoMSstatsFormat(openms_raw, use_log_file = FALSE)
head(openms_imported)

```

---

OpenSWATHtoMSstatsFormat

*Import OpenSWATH files*


---

## Description

Import OpenSWATH files

## Usage

```
OpenSWATHtoMSstatsFormat(
  input,
  annotation,
  filter_with_mscore = TRUE,
  mscore_cutoff = 0.01,
  useUniquePeptide = TRUE,
  removeFewMeasurements = TRUE,
  removeProtein_with1Feature = FALSE,
  summaryforMultipleRows = max,
  use_log_file = TRUE,
  append = FALSE,
  verbose = TRUE,
  log_file_path = NULL,
  ...
)
```

## Arguments

|                            |                                                                                                                                                        |
|----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| input                      | name of MSstats input report from OpenSWATH, which includes feature-level data.                                                                        |
| annotation                 | name of 'annotation.txt' data which includes Condition, BioReplicate, Run. Run should be the same as filename.                                         |
| filter_with_mscore         | TRUE(default) will filter out the features that have greater than mscore_cutoff in m_score column. Those features will be removed.                     |
| mscore_cutoff              | Cutoff for m_score. Default is 0.01.                                                                                                                   |
| useUniquePeptide           | TRUE (default) removes peptides that are assigned for more than one proteins. We assume to use unique peptide for each protein.                        |
| removeFewMeasurements      | TRUE (default) will remove the features that have 1 or 2 measurements across runs.                                                                     |
| removeProtein_with1Feature | TRUE will remove the proteins which have only 1 feature, which is the combination of peptide, precursor charge, fragment and charge. FALSE is default. |
| summaryforMultipleRows     | max(default) or sum - when there are multiple measurements for certain feature and certain run, use highest or sum of multiple intensities.            |
| 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. |
| ...           | additional parameters to <code>data.table::fread</code> .                                                                                                                                           |

**Value**

data.frame in the MSstats required format.

**Author(s)**

Meena Choi, Olga Vitek.

**Examples**

```
os_raw = system.file("tinytest/raw_data/OpenSWATH/openswath_input.csv",
                     package = "MSstatsConvert")
annot = system.file("tinytest/raw_data/OpenSWATH/annot_os.csv",
                   package = "MSstatsConvert")
os_raw = data.table::fread(os_raw)
annot = data.table::fread(annot)

os_imported = OpenSWATHtoMSstatsFormat(os_raw, annot, use_log_file = FALSE)
head(os_imported)
```

---

PDtoMSstatsFormat

---

*Import Proteome Discoverer files*


---

**Description**

Import Proteome Discoverer files

**Usage**

```
PDtoMSstatsFormat(
  input,
  annotation,
  useNumProteinsColumn = FALSE,
  useUniquePeptide = TRUE,
  summaryforMultipleRows = max,
  removeFewMeasurements = TRUE,
  removeOxidationMpeptides = FALSE,
  removeProtein_with1Peptide = FALSE,
  which.quantification = "Precursor.Area",
  which.proteinid = "Protein.Group.Accessions",
  which.sequence = "Sequence",
```

```

    use_log_file = TRUE,
    append = FALSE,
    verbose = TRUE,
    log_file_path = NULL,
    ...
)

```

## Arguments

|                                         |                                                                                                                                                                                                     |
|-----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>input</code>                      | PD report or a path to it.                                                                                                                                                                          |
| <code>annotation</code>                 | name of 'annotation.txt' or 'annotation.csv' data which includes Condition, BioReplicate, Run information. 'Run' will be matched with 'Spectrum.File'.                                              |
| <code>useNumProteinsColumn</code>       | TRUE removes peptides which have more than 1 in # Proteins column of PD output.                                                                                                                     |
| <code>useUniquePeptide</code>           | TRUE (default) removes peptides that are assigned for more than one proteins. We assume to use unique peptide for each protein.                                                                     |
| <code>summaryforMultipleRows</code>     | max(default) or sum - when there are multiple measurements for certain feature and certain run, use highest or sum of multiple intensities.                                                         |
| <code>removeFewMeasurements</code>      | TRUE (default) will remove the features that have 1 or 2 measurements across runs.                                                                                                                  |
| <code>removeOxidationMpeptides</code>   | TRUE will remove the peptides including 'oxidation (M)' in modification. FALSE is default.                                                                                                          |
| <code>removeProtein_with1Peptide</code> | TRUE will remove the proteins which have only 1 peptide and charge. FALSE is default.                                                                                                               |
| <code>which.quantification</code>       | Use 'Precursor.Area'(default) column for quantified intensities. 'Intensity' or 'Area' can be used instead.                                                                                         |
| <code>which.proteinid</code>            | Use 'Protein.Accessions'(default) column for protein name. 'Master.Protein.Accessions' can be used instead.                                                                                         |
| <code>which.sequence</code>             | Use 'Sequence'(default) column for peptide sequence. 'Annotated.Sequence' can be used instead.                                                                                                      |
| <code>use_log_file</code>               | logical. If TRUE, information about data processing will be saved to a file.                                                                                                                        |
| <code>append</code>                     | logical. If TRUE, information about data processing will be added to an existing log file.                                                                                                          |
| <code>verbose</code>                    | logical. If TRUE, information about data processing will be printed to the console.                                                                                                                 |
| <code>log_file_path</code>              | 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. |
| <code>...</code>                        | additional parameters to <code>data.table::fread</code> .                                                                                                                                           |

## Value

data.frame in the MSstats required format.

**Author(s)**

Meena Choi, Olga Vitek

**Examples**

```
pd_raw = system.file("tinytest/raw_data/PD/pd_input.csv",
                     package = "MSstatsConvert")
annot = system.file("tinytest/raw_data/PD/annot_pd.csv",
                   package = "MSstatsConvert")
pd_raw = data.table::fread(pd_raw)
annot = data.table::fread(annot)

pd_imported = PDtoMSstatsFormat(pd_raw, annot, use_log_file = FALSE)
head(pd_imported)
```

---

ProgenesistoMSstatsFormat

*Import Progenesis files*

---

**Description**

Import Progenesis files

**Usage**

```
ProgenesistoMSstatsFormat(
  input,
  annotation,
  useUniquePeptide = TRUE,
  summaryforMultipleRows = max,
  removeFewMeasurements = TRUE,
  removeOxidationMpeptides = FALSE,
  removeProtein_with1Peptide = FALSE,
  use_log_file = TRUE,
  append = FALSE,
  verbose = TRUE,
  log_file_path = NULL,
  ...
)
```

**Arguments**

|                  |                                                                                                                                                                          |
|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| input            | name of Progenesis output, which is wide-format. 'Accession', 'Sequence', 'Modification', 'Charge' and one column for each run are required.                             |
| annotation       | name of 'annotation.txt' or 'annotation.csv' data which includes Condition, BioReplicate, Run information. It will be matched with the column name of input for MS runs. |
| useUniquePeptide | TRUE (default) removes peptides that are assigned for more than one proteins. We assume to use unique peptide for each protein.                                          |



|                            |                                                                                                                                                                                                     |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| summaryforMultipleRows     | max(default) or sum - when there are multiple measurements for certain feature and certain run, use highest or sum of multiple intensities.                                                         |
| removeFewMeasurements      | TRUE (default) will remove the features that have 1 or 2 measurements across runs.                                                                                                                  |
| removeOxidationMpeptides   | TRUE will remove the peptides including 'oxidation (M)' in modification. FALSE is default.                                                                                                          |
| removeProtein_with1Peptide | TRUE will remove the proteins which have only 1 peptide and charge. FALSE is default.                                                                                                               |
| 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. |
| ...                        | additional parameters to <code>data.table::fread</code> .                                                                                                                                           |

**Value**

data.frame in the MSstats required format.

**Author(s)**

Meena Choi, Olga Vitek, Ulrich Omasits

**Examples**

```
progenesis_raw = system.file("tinytest/raw_data/Progenesis/progenesis_input.csv",
                             package = "MSstatsConvert")
annot = system.file("tinytest/raw_data/Progenesis/progenesis_annot.csv",
                   package = "MSstatsConvert")
progenesis_raw = data.table::fread(progenesis_raw)
annot = data.table::fread(annot)

progenesis_imported = ProgenesisToMSstatsFormat(progenesis_raw, annot,
                                                use_log_file = FALSE)
head(progenesis_imported)
```

---

ProteinProspectortoMSstatsTMTFormat

*Generate MSstatsTMT required input format from Protein Prospector output*

---

## Description

Generate MSstatsTMT required input format from Protein Prospector output

## Usage

```
ProteinProspectortoMSstatsTMTFormat(
  input,
  annotation,
  useUniquePeptide = TRUE,
  removeFewMeasurements = TRUE,
  removeProtein_with1Feature = FALSE,
  summaryforMultipleRows = sum,
  use_log_file = TRUE,
  append = FALSE,
  verbose = TRUE,
  log_file_path = NULL
)
```

## Arguments

|                            |                                                                                                                                                                                                     |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| input                      | Input txt peptide report file from Protein Prospector with "Keep Replicates", "Mods in Peptide", and "Protein Mods" options selected.                                                               |
| annotation                 | data frame which contains column Run, Fraction, TechRepMixture, Mixture, Channel, BioReplicate, Condition.                                                                                          |
| useUniquePeptide           | TRUE (default) removes peptides that are assigned for more than one proteins. We assume to use unique peptide for each protein.                                                                     |
| removeFewMeasurements      | TRUE (default) will remove the features that have 1 or 2 measurements across runs.                                                                                                                  |
| removeProtein_with1Feature | TRUE will remove the proteins which have only 1 feature, which is the combination of peptide, precursor charge, fragment and charge. FALSE is default.                                              |
| summaryforMultipleRows     | max(default) or sum - when there are multiple measurements for certain feature and certain run, use highest or sum of multiple intensities.                                                         |
| 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 wil 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. |

## Value

data.frame of class "MSstatsTMT"

**Examples**

```
input = system.file("tinytest/raw_data/ProteinProspector/Prospector_TotalTMT.txt",
  package = "MSstatsConvert")
input = data.table::fread(input)
annot = system.file("tinytest/raw_data/ProteinProspector/Annotation.csv",
  package = "MSstatsConvert")
annot = data.table::fread(annot)
output <- ProteinProspectortoMSstatsTMTFormat(input, annot)
head(output)
```

---

SkylinetoMSstatsFormat

*Import Skyline files*


---

**Description**

Import Skyline files

**Usage**

```
SkylinetoMSstatsFormat(
  input,
  annotation = NULL,
  removeiRT = TRUE,
  filter_with_Qvalue = TRUE,
  qvalue_cutoff = 0.01,
  useUniquePeptide = TRUE,
  removeFewMeasurements = TRUE,
  removeOxidationMpeptides = FALSE,
  removeProtein_with1Feature = FALSE,
  use_log_file = TRUE,
  append = FALSE,
  verbose = TRUE,
  log_file_path = NULL,
  ...
)
```

**Arguments**

|                    |                                                                                                                                                                                                                                      |
|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| input              | name of MSstats input report from Skyline, which includes feature-level data.                                                                                                                                                        |
| annotation         | name of 'annotation.txt' data which includes Condition, BioReplicate, Run. If annotation is already complete in Skyline, use annotation=NULL (default). It will use the annotation information from input.                           |
| removeiRT          | TRUE (default) will remove the proteins or peptides which are labeled 'iRT' in 'StandardType' column. FALSE will keep them.                                                                                                          |
| filter_with_Qvalue | TRUE(default) will filter out the intensities that have greater than qvalue_cutoff in DetectionQValue column. Those intensities will be replaced with zero and will be considered as censored missing values for imputation purpose. |

|                            |                                                                                                                                                                                                     |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| qvalue_cutoff              | Cutoff for DetectionQValue. default is 0.01.                                                                                                                                                        |
| useUniquePeptide           | TRUE (default) removes peptides that are assigned for more than one proteins. We assume to use unique peptide for each protein.                                                                     |
| removeFewMeasurements      | TRUE (default) will remove the features that have 1 or 2 measurements across runs.                                                                                                                  |
| removeOxidationMpeptides   | TRUE will remove the peptides including 'oxidation (M)' in modification. FALSE is default.                                                                                                          |
| removeProtein_with1Feature | TRUE will remove the proteins which have only 1 feature, which is the combination of peptide, precursor charge, fragment and charge. FALSE is default.                                              |
| 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. |
| ...                        | additional parameters to <code>data.table::fread</code> .                                                                                                                                           |

**Value**

data.frame in the MSstats required format.

**Author(s)**

Meena Choi, Olga Vitek

**Examples**

```
skyline_raw = system.file("tinytest/raw_data/Skyline/skyline_input.csv",
                          package = "MSstatsConvert")
skyline_raw = data.table::fread(skyline_raw)
skyline_imported = SkylinetoMSstatsFormat(skyline_raw)
head(skyline_imported)
```

---

SpectronauttoMSstatsFormat

*Import Spectronaut files*

---

**Description**

Import Spectronaut files

**Usage**

```

SpectronauttoMSstatsFormat(
  input,
  annotation = NULL,
  intensity = "PeakArea",
  filter_with_Qvalue = FALSE,
  qvalue_cutoff = 0.01,
  useUniquePeptide = TRUE,
  removeFewMeasurements = TRUE,
  removeProtein_with1Feature = FALSE,
  summaryforMultipleRows = max,
  use_log_file = TRUE,
  append = FALSE,
  verbose = TRUE,
  log_file_path = NULL,
  ...
)

```

**Arguments**

|                            |                                                                                                                                                                                                                                                                                                 |
|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| input                      | name of Spectronaut output, which is long-format. ProteinName, PeptideSequence, PrecursorCharge, FragmentIon, ProductCharge, IsotopeLabelType, Condition, BioReplicate, Run, Intensity, F.ExcludedFromQuantification are required. Rows with F.ExcludedFromQuantification=True will be removed. |
| annotation                 | name of 'annotation.txt' data which includes Condition, BioReplicate, Run. If annotation is already complete in Spectronaut, use annotation=NULL (default). It will use the annotation information from input.                                                                                  |
| intensity                  | 'PeakArea'(default) uses not normalized peak area. 'NormalizedPeakArea' uses peak area normalized by Spectronaut.                                                                                                                                                                               |
| filter_with_Qvalue         | FALSE(default) will not perform any filtering. TRUE will filter out the intensities that have greater than qvalue_cutoff in EG.Qvalue column. Those intensities will be replaced with zero and will be considered as censored missing values for imputation purpose.                            |
| qvalue_cutoff              | Cutoff for EG.Qvalue. default is 0.01.                                                                                                                                                                                                                                                          |
| useUniquePeptide           | TRUE (default) removes peptides that are assigned for more than one proteins. We assume to use unique peptide for each protein.                                                                                                                                                                 |
| removeFewMeasurements      | TRUE (default) will remove the features that have 1 or 2 measurements across runs.                                                                                                                                                                                                              |
| removeProtein_with1Feature | TRUE will remove the proteins which have only 1 feature, which is the combination of peptide, precursor charge, fragment and charge. FALSE is default.                                                                                                                                          |
| summaryforMultipleRows     | max(default) or sum - when there are multiple measurements for certain feature and certain run, use highest or sum of multiple intensities.                                                                                                                                                     |
| 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.                                                                                                                                                                                                      |

|                            |                                                                                                                                                                                                                   |
|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>verbose</code>       | logical. If TRUE, information about data processing wil be printed to the console.                                                                                                                                |
| <code>log_file_path</code> | 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 <code>append = TRUE</code> , has to be a valid path to a file. |
| <code>...</code>           | additional parameters to <code>data.table::fread</code> .                                                                                                                                                         |

**Value**

data.frame in the MSstats required format.

**Author(s)**

Meena Choi, Olga Vitek

**Examples**

```
spectronaut_raw = system.file("tinytest/raw_data/Spectronaut/spectronaut_input.csv",  
                             package = "MSstatsConvert")  
spectronaut_raw = data.table::fread(spectronaut_raw)  
spectronaut_imported = SpectronauttoMSstatsFormat(spectronaut_raw, use_log_file = FALSE)  
head(spectronaut_imported)
```

# Index

## \* internal

- .MSstatsFormat, 33
- .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
- .cleanRawMetamorpheus, 10
- .cleanRawOpenMS, 11
- .cleanRawOpenSWATH, 11
- .cleanRawPDMSstats, 12
- .cleanRawPDTMT, 13
- .cleanRawPhilosopher, 14
- .cleanRawProgenesis, 15
- .cleanRawSkyline, 15
- .cleanRawSpectroMineTMT, 16
- .cleanRawSpectronaut, 16
- .countCommonFeatures, 17
- .fillValues, 17
- .filterByPattern, 18
- .filterByScore, 18
- .filterExact, 19
- .filterFewMeasurements, 20
- .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, 30
- .makeBalancedDesign, 31
- .makeExactFilterMessage, 31
- .makeScoreFilterMessage, 32
- .mergeAnnotation, 32
- .nullAppender, 33
- .onLoad, 34
- .removeOverlappingFeatures, 34
- .removeSharedPeptides, 35
- .selectMSstatsColumns, 35
- .sharedParametersAmongConverters, 36
- .standardizeColnames, 37
- .summarizeMultipleMeasurements, 37
- .summarizeMultiplePSMs, 38
- getDataType, 44
- MSstatsInputFiles-class, 54
- .MSstatsFormat, 33
- .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
- .cleanRawMetamorpheus, 10
- .cleanRawOpenMS, 11
- .cleanRawOpenSWATH, 11
- .cleanRawPD, 12

- .cleanRawPDMSstats, 12
- .cleanRawPDTMT, 13
- .cleanRawPhilosopher, 14
- .cleanRawProgenesis, 15
- .cleanRawSkyline, 15
- .cleanRawSpectroMineTMT, 16
- .cleanRawSpectronaut, 16
- .countCommonFeatures, 17
- .fillValues, 17
- .filterByPattern, 18
- .filterByScore, 18
- .filterExact, 19
- .filterFewMeasurements, 20
- .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, 30
- .makeBalancedDesign, 31
- .makeExactFilterMessage, 31
- .makeScoreFilterMessage, 32
- .mergeAnnotation, 32
- .nullAppender, 33
- .onLoad, 34
- .removeOverlappingFeatures, 34
- .removeSharedPeptides, 35
- .selectMSstatsColumns, 35
- .sharedParametersAmongConverters, 36
- .standardizeColnames, 37
- .summarizeMultipleMeasurements, 37
- .summarizeMultiplePSMs, 38
- as.data.frame.MSstatsValidated, 38
- as.data.table.MSstatsValidated, 39
- DIANNtoMSstatsFormat, 39
- DIAUmpiretoMSstatsFormat, 41
- FragPipeToMSstatsFormat, 43
- getDataType, 44
- getDataType,MSstatsInputFiles-method  
(getDataType), 44
- getInputFile, 45
- getInputFile,MSstatsInputFiles-method  
(getInputFile), 45
- getInputFile,MSstatsPhilosopherFiles-method  
(getInputFile), 45
- MaxQtoMSstatsFormat, 45
- MetamorpheusToMSstatsFormat, 47
- MSstatsBalancedDesign, 49, 53
- MSstatsClean, 50, 53
- MSstatsClean,MSstatsDIANNFiles-method  
(MSstatsClean), 50
- MSstatsClean,MSstatsDIAUmpireFiles-method  
(MSstatsClean), 50
- MSstatsClean,MSstatsMaxQuantFiles-method  
(MSstatsClean), 50
- MSstatsClean,MSstatsMetamorpheusFiles-method  
(MSstatsClean), 50
- MSstatsClean,MSstatsOpenMSFiles-method  
(MSstatsClean), 50
- MSstatsClean,MSstatsOpenSWATHFiles-method  
(MSstatsClean), 50
- MSstatsClean,MSstatsPhilosopherFiles-method  
(MSstatsClean), 50
- MSstatsClean,MSstatsProgenesisFiles-method  
(MSstatsClean), 50
- MSstatsClean,MSstatsProteinProspectorFiles-method  
(MSstatsClean), 50
- MSstatsClean,MSstatsProteomeDiscovererFiles-method  
(MSstatsClean), 50
- MSstatsClean,MSstatsSkylineFiles-method  
(MSstatsClean), 50
- MSstatsClean,MSstatsSpectroMineFiles-method  
(MSstatsClean), 50
- MSstatsClean,MSstatsSpectronautFiles-method  
(MSstatsClean), 50
- MSstatsConvert, 53
- MSstatsDIANNFiles-class  
(MSstatsInputFiles-class), 54
- MSstatsDIAUmpireFiles-class  
(MSstatsInputFiles-class), 54
- MSstatsFragPipeFiles-class  
(MSstatsInputFiles-class), 54
- MSstatsImport, 53, 53
- MSstatsInputFiles-class, 54
- MSstatsLogsSettings, 53, 55
- MSstatsMakeAnnotation, 56



MSstatsMaxQuantFiles-class  
    (MSstatsInputFiles-class), [54](#)  
MSstatsMetamorpheusFiles-class  
    (MSstatsInputFiles-class), [54](#)  
MSstatsOpenMSFiles-class  
    (MSstatsInputFiles-class), [54](#)  
MSstatsOpenSWATHFiles-class  
    (MSstatsInputFiles-class), [54](#)  
MSstatsPhilosopherFiles-class  
    (MSstatsInputFiles-class), [54](#)  
MSstatsPreprocess, [53](#), [57](#)  
MSstatsProgenesisFiles-class  
    (MSstatsInputFiles-class), [54](#)  
MSstatsProteinProspectorFiles-class  
    (MSstatsInputFiles-class), [54](#)  
MSstatsProteomeDiscovererFiles-class  
    (MSstatsInputFiles-class), [54](#)  
MSstatsSaveSessionInfo, [59](#)  
MSstatsSkylineFiles-class  
    (MSstatsInputFiles-class), [54](#)  
MSstatsSpectroMineFiles-class  
    (MSstatsInputFiles-class), [54](#)  
MSstatsSpectronautFiles-class  
    (MSstatsInputFiles-class), [54](#)  
  
OpenMStoMSstatsFormat, [59](#)  
OpenSWATHtoMSstatsFormat, [61](#)  
  
PDtoMSstatsFormat, [62](#)  
ProgenesisistoMSstatsFormat, [64](#)  
ProteinProspectortoMSstatsTMTFormat,  
    [65](#)  
  
SkylinettoMSstatsFormat, [67](#)  
SpectronauttoMSstatsFormat, [68](#)