Package 'MAFFIN'

March 22, 2022

Type Package

Title Integrated Sample Normalization by MAFFIN Algorithm

Version 1.0.0

Author Huaxu Yu

Maintainer <hxyu@chem.ubc.ca>

Description MAFFIN sample normalization and evaluation functions.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

URL https://github.com/Waddlessss/MAFFIN

Depends R (>= 3.10)

Imports polynom,
stringr,
preprocessCore

R topics documented:

EvaPRMAD *Calculate pooled RMAD for normalization evaluation.*

Description

Calculate pooled relative median absolute deviation for each metabolic feature.

Usage

```
EvaPRMAD(FeatureTable, GroupNames, SampleInCol = TRUE, output = FALSE)
```

Arguments

FeatureTable	Data frame with features in row and samples in column (default).
GroupNames	A character vector indicating the names of each group.
SampleInCol	TRUE if samples are in column. FALSE if samples are in row.
output	TRUE will output the result table in current working directory

Details

FeatureTable contains measured signal intensities of metabolic features, with features in row and samples in column (default). The column names should be sample names, and the first row should be sample group names (e.g. control, case).
The first column should be unique feature identifiers. An example of FeatureTable is provided as TestingData in this package.

Value

This function will return a vector that contains the calculated PRMADs for all features.

References

Yu, Huaxu, and Tao Huan. "MAFFIN: Metabolomics Sample Normalization Using Maximal Density Fold Change with High-Quality Metabolic Features and Corrected Signal Intensities." *bioRxiv* (2021).

Examples

```
prmad = EvaPRMAD(TestingData, GroupNames=c("HY", "SX", "SW", "YC"))
```

EvaPRSD *Calculate pooled RSD for normalization evaluation.*

Description

Calculate pooled relative standard deviation for each metabolic feature.

Usage

```
EvaPRSD(FeatureTable, GroupNames, SampleInCol = TRUE, output = FALSE)
```

Arguments

FeatureTable	Data frame with features in row and samples in column (default).
GroupNames	A character vector indicating the names of each group.
SampleInCol	TRUE if samples are in column. FALSE if samples are in row.
output	TRUE will output the result table in current working directory

Details

FeatureTable contains measured signal intensities of metabolic features, with features in row and samples in column (default). The column names should be sample names, and the first row should be sample group names (e.g. control, case).
The first column should be unique feature identifiers. An example of FeatureTable is provided as TestingData in this package.

Value

This function will return a vector that contains the calculated PRSDs for all features.

References

Yu, Huaxu, and Tao Huan. "MAFFIN: Metabolomics Sample Normalization Using Maximal Density Fold Change with High-Quality Metabolic Features and Corrected Signal Intensities." *bioRxiv* (2021).

Examples

```
prsd = EvaPRMAD(TestingData, GroupNames=c("HY", "SX", "SW", "YC"))
```

`FeatureSelection` *High-quality feature selection*

Description

Select high-quality features for quantitative analysis.

Usage

```
FeatureSelection(
  FeatureTable,
  BlankFilter = 2,
  RtRange = c(0, 100),
  QCRSD = 0.25,
  SQCcor = 0.9,
  IntThreshold = 0,
  SampleInCol = TRUE,
  output = FALSE
)
```

Arguments

`FeatureTable`	Data frame with features in row and samples in column (default).
`BlankFilter`	A numeric value. High-quality when mean(sample intensities) > mean(blank intensities) * `BlankFilter`
`RtRange`	A numeric vector indicating the range of the defined retention time window, in minute.
`QCRSD`	A numeric value indicating the relative standard deviation threshold for QC samples.
`SQCcor`	A numeric value indicating the Pearson's correlation threshold for serial QC samples (recommend: 0.8-0.9).
`IntThreshold`	A numeric value indicating the feature intensity threshold. Feature is detected when its intensity larger than this value.
`SampleInCol`	`TRUE` if samples are in column. `FALSE` if samples are in row.
`output`	`TRUE` will output the result table in current working directory

Details

`FeatureTable` contains measured signal intensities of metabolic features, with features in row and samples in column (default). The column names should be sample names, and the first row should be sample group names (e.g. control, case).
The first column should be unique feature identifiers. For group names, please use:
"RT" for retention time column;
"QC" for quality control samples between real samples (normal QC samples);
"blank" for blank samples;
"SQC_###" for serial QC samples with a certain loading amount. For example, SQC_1.0 means a serial QC sample with injection volume of 1.0 uL.
An example of `FeatureTable` is provided as `TestingData` in this package.

Value

This function will return the original data frame with an extra column named "Quality" to indicate the feature quality.

References

Yu, Huaxu, and Tao Huan. "MAFFIN: Metabolomics Sample Normalization Using Maximal Density Fold Change with High-Quality Metabolic Features and Corrected Signal Intensities." *bioRxiv* (2021).

Examples

```
selectedTable = FeatureSelection(TestingData)
```

IntCorrection *Intensity correction using serial QC samples*

Description

Correct MS signal intensities using serial QC samples

Usage

```
IntCorrection(
  FeatureTable,
  IntThreshold = 0,
  LR_QC_points = 5,
  QR_QC_points = 7,
  SQCcor = 0.9,
  SampleInCol = TRUE,
  output = FALSE
)
```

Arguments

FeatureTable	Data frame with features in row and samples in column (default).
IntThreshold	A numeric value indicating the feature intensity threshold. Feature is detected when its intensity larger than this value.
LR_QC_points	Minimum serial QC data points for quadratic regression.
QR_QC_points	Minimum serial QC data points for cubic regression.
SQCcor	Pearson's correlation threshold for serial QC samples (recommend: 0.8-0.9).
SampleInCol	TRUE if samples are in column. FALSE if samples are in row.
output	TRUE will output the result table in current working directory.

Details

FeatureTable contains measured signal intensities of metabolic features, with features in row and samples in column (default). The column names should be sample names, and the first row should be sample group names (e.g. control, case).
The first column should be unique feature identifiers. For group names, please use:
"QC" for quality control samples between real samples (normal QC samples);
"SQC_###" for serial QC samples with a certain loading amount. For example, SQC_1.0 means a serial QC sample with injection volume of 1.0 uL.
Please note, group names of real biological samples cannot be "RT" and "blank".
An example of FeatureTable is provided as TestingData in this package.

Value

This function will return the original feature table with corrected intensities.

References

Yu, Huaxu, and Tao Huan. "MAFFIN: Metabolomics Sample Normalization Using Maximal Density Fold Change with High-Quality Metabolic Features and Corrected Signal Intensities." *bioRxiv* (2021).

Examples

```
intCorrectedTable = IntCorrection(TestingData)
```

MAFFINNorm *MAFFFIN normalization*

Description

Perform sample normalization using MAFFIN algorithm. MAFFIN algorithm consists of three modules: high-quality feature selection, MS signal intensity correction, and maximal density fold change normalization.

Usage

```
MAFFINNorm(
  FeatureTable,
  BlankFilter = 2,
  RtRange = c(0, 100),
  QCRSD = 0.25,
  SQCcor = 0.9,
  IntThreshold = 0,
  LR_QC_points = 5,
  QR_QC_points = 7,
  SampleInCol = TRUE,
  output = FALSE
)
```

Arguments

FeatureTable	Data frame with features in row and samples in column (default).
BlankFilter	A numeric value. High-quality when mean(sample intensities) > mean(blank intensities) * BlankFilter.
RtRange	A numeric vector indicating the range of the defined retention time window, in minute.
QCRSD	A numeric value indicating the relative standard deviation threshold for QC samples.
SQCcor	A numeric value indicating the Pearson's correlation threshold for serial QC samples (recommend: 0.8-0.9).
IntThreshold	A numeric value indicating the feature intensity threshold. Feature is detected when its intensity larger than this value.
LR_QC_points	Minimum serial QC data points for quadratic regression.
QR_QC_points	Minimum serial QC data points for cubic regression.
SampleInCol	TRUE if samples are in column. FALSE if samples are in row.
output	TRUE will output the result table in current working directory.

Details

FeatureTable contains measured signal intensities of metabolic features, with features in row and samples in column (default). The column names should be sample names, and the first row should be sample group names (e.g. control, case).
The first column should be unique feature identifiers. For group names, please use:
"RT" for retention time column;
"QC" for quality control samples between real samples (normal QC samples);
"blank" for blank samples;
"SQC_###" for serial QC samples with a certain loading amount. For example, SQC_1.0 means a serial QC sample with injection volume of 1.0 uL.
An example of FeatureTable is provided as TestingData in this package.

Value

This function will return a list that contains four items if RunEvaluation = TRUE: the normalized feature table, normalization factors, PRMAD of original data, and PRMAD of normalized data. The last two items will not be generated if RunEvaluation = FALSE

References

Yu, Huaxu, and Tao Huan. "MAFFIN: Metabolomics Sample Normalization Using Maximal Density Fold Change with High-Quality Metabolic Features and Corrected Signal Intensities." *bioRxiv* (2021).

Examples

```
MAFFINTable = MAFFINNorm(TestingData)
```

MDFCNorm *Maximal density fold change normalization*

Description

Sample normalization using maximal density fold change.

Usage

```
MDFCNorm(
  FeatureTable,
  IntThreshold = 0,
  SampleInCol = TRUE,
  output = FALSE,
  OutputNormFactors = FALSE,
  RunEvaluation = TRUE,
  bwOpt = NA
)
```

Arguments

FeatureTable	Data frame with features in row and samples in column (default).
IntThreshold	A numeric value indicating the feature intensity threshold. Feature is detected when its intensity larger than this value.
SampleInCol	TRUE if samples are in column. FALSE if samples are in row.
output	TRUE will output the result table in current working directory
OutputNormFactors	TRUE will print the normalization factors after normalization
RunEvaluation	TRUE will evaluate the normalization results by intragroup variation.
bwOpt	NA will automatically optimize the bandwidth. Use a numeric value to set the bandwidth and skip the optimization.

Details

FeatureTable contains measured or corrected signal intensities of metabolic features, with features in row and samples in column (default). The column names should be sample names, and the first row should be sample group names (e.g. control, case).
The first column should be unique feature identifiers. For group names, please use:
"RT" for retention time column;
"QC" for quality control samples between real samples (normal QC samples);
"blank" for blank samples;
"SQC_###" for serial QC samples with a certain loading amount. For example, SQC_1.0 means a serial QC sample with injection volume of 1.0 uL.
An example of FeatureTable is provided as TestingData in this package.

Value

This function will return a list that contains four items if RunEvaluation = TRUE: the normalized feature table, normalization factors, PRMAD of original data, and PRMAD of normalized data. The last two items will not be generated if RunEvaluation = FALSE

References

Yu, Huaxu, and Tao Huan. "MAFFIN: Metabolomics Sample Normalization Using Maximal Density Fold Change with High-Quality Metabolic Features and Corrected Signal Intensities." *bioRxiv* (2021).

Examples

```
MDFCNormedTable = MDFCNorm(TestingData)
```

`MedianNorm` *Normalization by median intensity*

Description

Sample normalization by median of MS signal intensity.

Usage

```
MedianNorm(
  FeatureTable,
  IntThreshold = 0,
  SampleInCol = TRUE,
  output = FALSE,
  OutputNormFactors = FALSE,
  RunEvaluation = TRUE
)
```

Arguments

`FeatureTable`	Data frame with features in row and samples in column (default).
`IntThreshold`	A numeric value indicating the feature intensity threshold. Feature is detected when its intensity larger than this value.
`SampleInCol`	`TRUE` if samples are in column. `FALSE` if samples are in row.
`output`	`TRUE` will output the result table in current working directory
`OutputNormFactors`	`TRUE` will print the normalization factors after normalization
`RunEvaluation`	`TRUE` will evaluate the normalization results using intragroup variation.

Details

`FeatureTable` contains measured or corrected signal intensities of metabolic features, with features in row and samples in column (default). The column names should be sample names, and the first row should be sample group names (e.g. control, case).
The first column should be unique feature identifiers. For group names, please do not use "blank", "RT", "QC", or "SQC_###" for real biological samples.
An example of `FeatureTable` is provided as `TestingData` in this package.

Value

This function will return a list that contains four items if `RunEvaluation = TRUE`: the normalized feature table, normalization factors, PRMAD of original data, and PRMAD of normalized data. The last two items will not be generated if `RunEvaluation = FALSE`

References

Yu, Huaxu, and Tao Huan. "MAFFIN: Metabolomics Sample Normalization Using Maximal Density Fold Change with High-Quality Metabolic Features and Corrected Signal Intensities." *bioRxiv* (2021).

Examples

```
MedianNormedTable = MedianNorm(TestingData)
```

`PQNNorm` *Probabilistic quotient normalization.*

Description

Sample normalization using median fold change.

Usage

```
PQNNorm(
  FeatureTable,
  IntThreshold = 0,
  SampleInCol = TRUE,
  output = FALSE,
  OutputNormFactors = FALSE,
  RunEvaluation = TRUE
)
```

Arguments

`FeatureTable`	Data frame with features in row and samples in column (default).
`IntThreshold`	A numeric value indicating the feature intensity threshold. Feature is detected when its intensity larger than this value.
`SampleInCol`	`TRUE` if samples are in column. `FALSE` if samples are in row.
`output`	`TRUE` will output the result table in current working directory
`OutputNormFactors`	`TRUE` will print the normalization factors after normalization
`RunEvaluation`	`TRUE` will evaluate the normalization results using intragroup variation.

Details

`FeatureTable` contains measured or corrected signal intensities of metabolic features, with features in row and samples in column (default). The column names should be sample names, and the first row should be sample group names (e.g. control, case).
The first column should be unique feature identifiers. For group names, please do not use "blank", "RT", "QC", or "SQC_###" for real biological samples.
An example of `FeatureTable` is provided as `TestingData` in this package.

Value

This function will return a list that contains four items if `RunEvaluation = TRUE`: the normalized feature table, normalization factors, PRMAD of original data, and PRMAD of normalized data. The last two items will not be generated if `RunEvaluation = FALSE`

References

Yu, Huaxu, and Tao Huan. "MAFFIN: Metabolomics Sample Normalization Using Maximal Density Fold Change with High-Quality Metabolic Features and Corrected Signal Intensities." *bioRxiv* (2021).
Dieterle, Frank, et al. "Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics." *Analytical chemistry* 78.13 (2006): 4281-4290.

Examples

```
PQNNormedTable = PQNNorm(TestingData)
```

`QuantileNorm` *Quantile normalization.*

Description

Quantile sample normalization.

Usage

```
QuantileNorm(
  FeatureTable,
  IntThreshold = 0,
  SampleInCol = TRUE,
  output = FALSE,
  OutputNormFactors = FALSE,
  RunEvaluation = TRUE
)
```

Arguments

`FeatureTable`	Data frame with features in row and samples in column (default).
`IntThreshold`	A numeric value indicating the feature intensity threshold. Feature is detected when its intensity larger than this value.
`SampleInCol`	`TRUE` if samples are in column. `FALSE` if samples are in row.

output	TRUE will output the result table in current working directory
OutputNormFactors	TRUE will print the normalization factors after normalization
RunEvaluation	TRUE will evaluate the normalization results using intragroup variation.

Details

FeatureTable contains measured or corrected signal intensities of metabolic features, with features in row and samples in column (default). The column names should be sample names, and the first row should be sample group names (e.g. control, case).
The first column should be unique feature identifiers. For group names, please do not use "blank", "RT", "QC", or "SQC_###" for real biological samples.
An example of FeatureTable is provided as TestingData in this package.

Value

This function will return a list that contains four items if RunEvaluation = TRUE: the normalized feature table, normalization factors, PRMAD of original data, and PRMAD of normalized data. The last two items will not be generated if RunEvaluation = FASLE

References

Yu, Huaxu, and Tao Huan. "MAFFIN: Metabolomics Sample Normalization Using Maximal Density Fold Change with High-Quality Metabolic Features and Corrected Signal Intensities." *bioRxiv* (2021).
Bolstad, Benjamin M., et al. "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias." *Bioinformatics* 19.2 (2003): 185-193.

Examples

```
QuantileNormedTable = QuantileNorm(TestingData)
```

SumNorm *Normalization by sum intensity*

Description

Sample normalization by total MS signal intensity.

Usage

```
SumNorm(
  FeatureTable,
  IntThreshold = 0,
  SampleInCol = TRUE,
  output = FALSE,
  OutputNormFactors = FALSE,
  RunEvaluation = TRUE
)
```

Arguments

FeatureTable	Data frame with features in row and samples in column (default).
IntThreshold	A numeric value indicating the feature intensity threshold. Feature is detected when its intensity larger than this value.
SampleInCol	TRUE if samples are in column. FALSE if samples are in row.
output	TRUE will output the result table in current working directory
OutputNormFactors	TRUE will print the normalization factors after normalization
RunEvaluation	TRUE will evaluate the normalization results using intragroup variation.

Details

FeatureTable contains measured or corrected signal intensities of metabolic features, with features in row and samples in column (default). The column names should be sample names, and the first row should be sample group names (e.g. control, case).
The first column should be unique feature identifiers. For group names, please do not use "blank", "RT", "QC", or "SQC_###" for real biological samples.
An example of FeatureTable is provided as TestingData in this package.

Value

This function will return a list that contains four items if RunEvaluation = TRUE: the normalized feature table, normalization factors, PRMAD of original data, and PRMAD of normalized data. The last two items will not be generated if RunEvaluation = FALSE

References

Yu, Huaxu, and Tao Huan. "MAFFIN: Metabolomics Sample Normalization Using Maximal Density Fold Change with High-Quality Metabolic Features and Corrected Signal Intensities." *bioRxiv* (2021).

Examples

```
SumNormedTable = SumNorm(TestingData)
```

Index