

Package ‘pmd’

January 22, 2021

Type Package

Title Paired Mass Distance Analysis for GC/LC-MS Based Non-Targeted Analysis and Reactomics Analysis

Version 0.2.1

Date 2021-01-20

Maintainer Miao YU <yufreecas@gmail.com>

Description Paired mass distance (PMD) analysis proposed in Yu, Olkowicz and Pawliszyn (2018) <doi:10.1016/j.aca.2018.10.062> for gas/liquid chromatography–mass spectrometry (GC/LC-MS) based non-targeted analysis. PMD analysis including GlobalStd algorithm and structure/reaction directed analysis. GlobalStd algorithm could found independent peaks in m/z-retention time profiles based on retention time hierarchical cluster analysis and frequency analysis of paired mass distances within retention time groups. Structure directed analysis could be used to find potential relationship among those independent peaks in different retention time groups based on frequency of paired mass distances. Reactomics analysis could also be performed to build PMD network, assign sources and make biomarker reaction discovery. GUIs for PMD analysis is also included as 'shiny' applications.

URL <https://yufree.github.io/pmd/>

BugReports <https://github.com/yufree/pmd/issues>

License GPL-2

Encoding UTF-8

LazyData true

Suggests knitr, shiny, rmarkdown, MSnbase

VignetteBuilder knitr

biocViews

Depends R (>= 3.5.0)

Imports RColorBrewer, stats, graphics, utils, data.table, igraph, enviGCMS

RoxygenNote 7.1.1

NeedsCompilation no

Author Miao YU [aut, cre] (<<https://orcid.org/0000-0002-2804-6014>>)

Repository CRAN

Date/Publication 2021-01-21 23:20:08 UTC

R topics documented:

getcda	2
getchain	3
getcluster	4
getcorcluster	5
getms2pmd	5
getmspmd	6
getpaired	6
getpmd	7
getposneg	8
getrda	9
getreact	10
getsda	11
getstd	12
gettarget	13
globalstd	13
hmdb	14
kegrall	15
omics	15
pcasf	16
ploten	17
plotpaired	17
plotrtg	18
plotsda	18
plotstd	19
plotstdrt	19
plotstdsda	20
pmdanno	21
runPMD	21
runPMDnet	21
sda	22
spmein vivo	22
Index	23

getcda	<i>Perform correlation directed analysis for peaks list.</i>
--------	--

Description

Perform correlation directed analysis for peaks list.

Usage

```
getcda(list, corcutoff = 0.9, rtcutoff = 10, accuracy = 4)
```

Arguments

list	a list with mzrt profile
corcutoff	cutoff of the correlation coefficient, default NULL
rtcutoff	cutoff of the distances in retention time hierarchical clustering analysis, default 10
accuracy	measured mass or mass to charge ratio in digits, default 4

Value

list with correlation directed analysis results

See Also

[getsda](#), [getrda](#)

Examples

```
data(spmein vivo)
cluster <- getcorcluster(spmein vivo)
cbp <- enviGMS::getfilter(cluster, rowindex = cluster$stdmassindex2)
cda <- getcda(cbp)
```

getchain

Get reaction chain for specific mass to charge ratio

Description

Get reaction chain for specific mass to charge ratio

Usage

```
getchain(
  list,
  diff,
  mass,
  digits = 2,
  accuracy = 4,
  rtcutoff = 10,
  corcutoff = 0.6,
  ppm = 25
)
```

Arguments

list	a list with mzrt profile
diff	paired mass distance(s) of interests
mass	a specific mass for known compound or a vector of masses. You could also input formula for certain compounds
digits	mass or mass to charge ratio accuracy for pmd, default 2
accuracy	measured mass or mass to charge ratio in digits, default 4
rtcutoff	cutoff of the distances in retention time hierarchical clustering analysis, default 10
corcutoff	cutoff of the correlation coefficient, default NULL
ppm	all the peaks within this mass accuracy as seed mass or formula

Value

a list with mzrt profile and reaction chain dataframe

Examples

```
data(spmein vivo)
# check metabolites of C18H39NO
pmd <- getchain(spmein vivo,diff = c(2.02,14.02,15.99),mass = 286.3101)
```

getcluster

Get Pseudo-Spectrum as peaks cluster based on pmd analysis.

Description

Get Pseudo-Spectrum as peaks cluster based on pmd analysis.

Usage

```
getcluster(list, corcutoff = NULL, accuracy = 4)
```

Arguments

list	a list from getstd function
corcutoff	cutoff of the correlation coefficient, default NULL
accuracy	measured mass or mass to charge ratio in digits, default 4

Value

list with Pseudo-Spectrum index

See Also

[getpaired](#),[getstd](#),[plotstd](#)

Examples

```
data(spmein vivo)
re <- getpaired(spmein vivo)
re <- getstd(re)
cluster <- getcluster(re)
```

getcorcluster	<i>Get Pseudo-Spectrum as peaks cluster based on correlation analysis.</i>
---------------	--

Description

Get Pseudo-Spectrum as peaks cluster based on correlation analysis.

Usage

```
getcorcluster(list, corcutoff = 0.9, rtcutoff = 10, accuracy = 4)
```

Arguments

list	a list with peaks intensity
corcutoff	cutoff of the correlation coefficient, default 0.9
rtcutoff	cutoff of the distances in cluster, default 10
accuracy	measured mass or mass to charge ratio in digits, default 4

Value

list with Pseudo-Spectrum index

Examples

```
data(spmein vivo)
cluster <- getcorcluster(spmein vivo)
```

getms2pmd	<i>read in MSP file as list for ms/ms annotation</i>
-----------	--

Description

read in MSP file as list for ms/ms annotation

Usage

```
getms2pmd(file, digits = 2, icf = 10)
```

Arguments

file the path to your MSP file
 digits mass or mass to charge ratio accuracy for pmd, default 2
 icf intensity cutoff, default 10 percentage

Value

list a list with MSP information for MS/MS annotation

getmspmd *read in MSP file as list for EI-MS annotation*

Description

read in MSP file as list for EI-MS annotation

Usage

```
getmspmd(file, digits = 2, icf = 10)
```

Arguments

file the path to your MSP file
 digits mass or mass to charge ratio accuracy for pmd, default 0
 icf intensity cutoff, default 10 percentage

Value

list a list with MSP information for EI-MS annotation

getpaired *Filter ions/peaks based on retention time hierarchical clustering, paired mass distances(PMD) and PMD frequency analysis.*

Description

Filter ions/peaks based on retention time hierarchical clustering, paired mass distances(PMD) and PMD frequency analysis.

Usage

```
getpaired(list, rtcutoff = 10, ng = NULL, digits = 2, accuracy = 4)
```

Arguments

list	a list with mzrt profile
rtcutoff	cutoff of the distances in retention time hierarchical clustering analysis, default 10
ng	cutoff of global PMD's retention time group numbers, If ng = NULL, 20 percent of RT cluster will be used as ng, default NULL.
digits	mass or mass to charge ratio accuracy for pmd, default 2
accuracy	measured mass or mass to charge ratio in digits, default 4

Value

list with tentative isotope, multi-chargers, adducts, and neutral loss peaks' index, retention time clusters.

See Also

[getstd](#), [getsda](#), [plotpaired](#)

Examples

```
data(spmein vivo)
pmd <- getpaired(spmein vivo)
```

getpmd

Get pmd for specific reaction

Description

Get pmd for specific reaction

Usage

```
getpmd(list, pmd, rtcutoff = 10, corcutoff = NULL, digits = 2, accuracy = 4)
```

Arguments

list	a list with mzrt profile
pmd	a specific paired mass distance
rtcutoff	cutoff of the distances in retention time hierarchical clustering analysis, default 10
corcutoff	cutoff of the correlation coefficient, default NULL
digits	mass or mass to charge ratio accuracy for pmd, default 2
accuracy	measured mass or mass to charge ratio in digits, default 4

Value

list with paired peaks for specific pmd.

See Also

[getpaired](#), [getstd](#), [getsda](#), [getrda](#)

Examples

```
data(spmein vivo)
pmd <- getpmd(spmein vivo, pmd=15.99)
```

getposneg

Link pos mode peak list with neg mode peak list by pmd.

Description

Link pos mode peak list with neg mode peak list by pmd.

Usage

```
getposneg(pos, neg, pmd = 2.02, digits = 2)
```

Arguments

pos	a list with mzrt profile collected from positive mode.
neg	a list with mzrt profile collected from negative mode.
pmd	numeric or numeric vector
digits	mass or mass to charge ratio accuracy for pmd, default 2

Value

dataframe with filtered positive and negative peak list

getrda	<i>Perform structure/reaction directed analysis for mass only.</i>
--------	--

Description

Perform structure/reaction directed analysis for mass only.

Usage

```
getrda(mz, freqcutoff = 10, digits = 3, top = 20, formula = NULL)
```

Arguments

mz	numeric vector for independent mass or mass to charge ratio. Mass to charge ratio from GlobalStd algorithm is suggested. Isomers would be excluded automatically
freqcutoff	pmd frequency cutoff for structures or reactions, default 10
digits	mass or mass to charge ratio accuracy for pmd, default 3
top	top n pmd frequency cutoff when the freqcutoff is too small for large data set
formula	vector for formula when you don't have mass or mass to charge ratio data

Value

logical matrix with row as the same order of mz or formula and column as high frequency pmd group

See Also

[getsda](#)

Examples

```
data(spmein vivo)
pmd <- getpaired(spmein vivo)
std <- getstd(pmd)
sda <- getrda(spmein vivo$mz[std$stdmassindex])
```

 getreact

Get quantitative paired peaks list for specific reaction/pmd

Description

Get quantitative paired peaks list for specific reaction/pmd

Usage

```
getreact(
  list,
  pmd,
  rtcutoff = 10,
  digits = 2,
  accuracy = 4,
  ratio cv = 30,
  outlier = F,
  ...
)
```

Arguments

list	a list with mzrt profile and data
pmd	a specific paired mass distances
rtcutoff	cutoff of the distances in retention time hierarchical clustering analysis, default 10
digits	mass or mass to charge ratio accuracy for pmd, default 2
accuracy	measured mass or mass to charge ratio in digits, default 4
ratio cv	ratio cv cutoff for quantitative paired peaks, default 30
outlier	logical, if true, outlier of ratio will be removed, default False.
...	other parameters for getpmd

Value

list with quantitative paired peaks.

See Also

[getpaired](#), [getstd](#), [getsda](#), [getrda](#), [getpmd](#),

Examples

```
data(spmein vivo)
pmd <- getreact(spmein vivo, pmd=15.99)
```

getsda	<i>Perform structure/reaction directed analysis for peaks list.</i>
--------	---

Description

Perform structure/reaction directed analysis for peaks list.

Usage

```
getsda(  
  list,  
  rtcutoff = 10,  
  corcutoff = NULL,  
  digits = 2,  
  accuracy = 4,  
  freqcutoff = NULL  
)
```

Arguments

<code>list</code>	a list with mzrt profile
<code>rtcutoff</code>	cutoff of the distances in retention time hierarchical clustering analysis, default 10
<code>corcutoff</code>	cutoff of the correlation coefficient, default NULL
<code>digits</code>	mass or mass to charge ratio accuracy for pmd, default 2
<code>accuracy</code>	measured mass or mass to charge ratio in digits, default 4
<code>freqcutoff</code>	pmd frequency cutoff for structures or reactions, default NULL. This cutoff will be found by PMD network analysis when it is NULL.

Value

list with tentative isotope, adducts, and neutral loss peaks' index, retention time clusters.

See Also

[getpaired](#), [getstd](#), [plotpaired](#)

Examples

```
data(spmein vivo)  
pmd <- getpaired(spmein vivo)  
std <- getstd(pmd)  
sda <- getsda(std)
```

getstd *Find the independent ions for each retention time hierarchical clustering based on PMD relationship within each retention time cluster and isotope and return the index of the std data for each retention time cluster.*

Description

Find the independent ions for each retention time hierarchical clustering based on PMD relationship within each retention time cluster and isotope and return the index of the std data for each retention time cluster.

Usage

```
getstd(list, corcutoff = NULL, digits = 2, accuracy = 4)
```

Arguments

list	a list from getpaired function
corcutoff	cutoff of the correlation coefficient, default NULL
digits	mass or mass to charge ratio accuracy for pmd, default 2
accuracy	measured mass or mass to charge ratio in digits, default 4

Value

list with std mass index

See Also

[getpaired](#), [getsda](#), [plotstd](#)

Examples

```
data(spmeinivo)
pmd <- getpaired(spmeinivo)
std <- getstd(pmd)
```

gettarget	<i>Get multiple injections index for selected retention time</i>
-----------	--

Description

Get multiple injections index for selected retention time

Usage

```
gettarget(rt, drt = 10, n = 6)
```

Arguments

rt	retention time vector for peaks in seconds
drt	retention time drift for targeted analysis in seconds, default 10.
n	max ions numbers within retention time drift windows

Value

index for each injection

Examples

```
data(spmeinivo)
pmd <- getpaired(spmeinivo)
std <- getstd(pmd)
index <- gettarget(std$rt[std$stdmassindex])
table(index)
```

globalstd	<i>GlobalStd algorithm with structure/reaction directed analysis</i>
-----------	--

Description

GlobalStd algorithm with structure/reaction directed analysis

Usage

```
globalstd(
  list,
  rtcutoff = 10,
  ng = NULL,
  corcutoff = NULL,
  digits = 2,
  accuracy = 4,
  freqcutoff = NULL,
  sda = FALSE
)
```

Arguments

<code>list</code>	a peaks list with mass to charge, retention time and intensity data
<code>rtcutoff</code>	cutoff of the distances in cluster, default 10
<code>ng</code>	cutoff of global PMD's retention time group numbers, If <code>ng = NULL</code> , 20 percent of RT cluster will be used as <code>ng</code> , default <code>NULL</code> .
<code>corcutoff</code>	cutoff of the correlation coefficient, default <code>NULL</code>
<code>digits</code>	mass or mass to charge ratio accuracy for <code>pmd</code> , default 2
<code>accuracy</code>	measured mass or mass to charge ratio in digits, default 4
<code>freqcutoff</code>	<code>pmd</code> frequency cutoff for structures or reactions, default <code>NULL</code> . This cutoff will be found by PMD network analysis when it is <code>NULL</code> .
<code>sda</code>	logical, option to perform structure/reaction directed analysis, default <code>FALSE</code> .

Value

list with GlobalStd algorithm processed data.

See Also

[getpaired](#), [getstd](#), [getsda](#), [plotstd](#), [plotstdsda](#), [plotstdrt](#)

Examples

```
data(spmein vivo)
re <- globalstd(spmein vivo)
```

<code>hmdb</code>	<i>A dataframe containing HMDB with unique accurate mass <code>pmd</code> with three digits frequency larger than 1 and accuracy percentage larger than 0.9.</i>
-------------------	--

Description

A dataframe containing HMDB with unique accurate mass `pmd` with three digits frequency larger than 1 and accuracy percentage larger than 0.9.

Usage

```
data(hmdb)
```

Format

A dataframe with atoms numbers of C, H, O, N, P, S

percentage accuracy of atom numbers prediction

pmd2 `pmd` with two digits

pmd `pmd` with three digits

keggrall	<i>A dataframe containing reaction related accurate mass pmd and related reaction formula with KEGG ID</i>
----------	--

Description

A dataframe containing reaction related accurate mass pmd and related reaction formula with KEGG ID

Usage

```
data(keggrall)
```

Format

A dataframe with KEGG reaction, their related pmd and atoms numbers of C, H, O, N, P, S

ID KEGG reaction ID

pmd pmd with three digits

omics	<i>A dataframe containing multiple reaction database ID and their related accurate mass pmd and related reactions</i>
-------	---

Description

A dataframe containing multiple reaction database ID and their related accurate mass pmd and related reactions

Usage

```
data(omics)
```

Format

A dataframe with reaction and their related pmd

KEGG KEGG reaction ID

RHEA_ID RHEA_ID

DIRECTION reaction direction

MASTER_ID master reaction RHEA ID

ec ec reaction ID

ecocyc ecocyc reaction ID

macie macie reaction ID

metacyc metacyc reaction ID
reactome reactome reaction ID
compounds reaction related compounds
pmd pmd with two digits
pmd2 pmd with three digits

pcasf

Compare matrices using PCA similarity factor

Description

Compare matrices using PCA similarity factor

Usage

```
pcasf(x, y, dim = NULL)
```

Arguments

x	Matrix with sample in column and features in row
y	Matrix is compared to x.
dim	number of retained dimensions in the comparison. Defaults to all.

Value

Ratio of projected variance to total variance

Author(s)

Edgar Zanella Alvarenga

References

Singhal, A. and Seborg, D. E. (2005), Clustering multivariate time-series data. *J. Chemometrics*, 19: 427-438. doi: 10.1002/cem.945

Examples

```
c1 <- matrix(rnorm(16),nrow=4)
c2 <- matrix(rnorm(16),nrow=4)
pcasf(c1, c2)
```

plotcn	<i>plot PMD KEGG network for certain compounds and output network average distance and degree</i>
--------	---

Description

plot PMD KEGG network for certain compounds and output network average distance and degree

Usage

```
plotcn(formula, name, pmd)
```

Arguments

formula	Chemical formula
name	Compound name
pmd	specific paired mass distances

Examples

```
plotcn('C6H12O6', 'Glucose', c(2.016, 14.016, 15.995))
```

plotpaired	<i>Plot the mass pairs and high frequency mass distances</i>
------------	--

Description

Plot the mass pairs and high frequency mass distances

Usage

```
plotpaired(list, index = NULL, ...)
```

Arguments

list	a list from getpaired function
index	index for PMD value
...	other parameters for plot function

See Also

[getpaired](#), [globalstd](#)

Examples

```
data(spmein vivo)
pmd <- getpaired(spmein vivo)
plotpaired(pmd)
```

plotrtg *Plot the retention time group*

Description

Plot the retention time group

Usage

```
plotrtg(list, ...)
```

Arguments

`list` a list from `getpaired` function
`...` other parameters for plot function

See Also

[getpaired](#), [globalstd](#)

Examples

```
data(spmein vivo)  
pmd <- getpaired(spmein vivo)  
plotrtg(pmd)
```

plotsda *Plot the specific structure directed analysis(SDA) groups*

Description

Plot the specific structure directed analysis(SDA) groups

Usage

```
plotsda(list, ...)
```

Arguments

`list` a list from `getpmd` function
`...` other parameters for plot function

See Also

[getstd](#), [globalstd](#), [plotstd](#), [plotpaired](#), [plotstdrt](#)

Examples

```
data(spmein vivo)
re <- getpmd(spmein vivo, pmd=78.9)
plotsda(re)
```

plotstd	<i>Plot the std mass from GlobalStd algorithm</i>
---------	---

Description

Plot the std mass from GlobalStd algorithm

Usage

```
plotstd(list)
```

Arguments

`list` a list from `getstd` function

See Also

[getstd](#), [globalstd](#)

Examples

```
data(spmein vivo)
pmd <- getpaired(spmein vivo)
std <- getstd(pmd)
plotstd(std)
```

plotstdrt	<i>Plot the std mass from GlobalStd algorithm in certain retention time groups</i>
-----------	--

Description

Plot the std mass from GlobalStd algorithm in certain retention time groups

Usage

```
plotstdrt(list, rtcluster, ...)
```

Arguments

`list` a list from `getstd` function
`rtcluster` retention time group index
`...` other parameters for plot function

See Also

[getstd](#), [globalstd](#), [plotstd](#), [plotpaired](#), [plotstdsda](#)

Examples

```
data(spmein vivo)
pmd <- getpaired(spmein vivo)
std <- getstd(pmd)
plotstdrt(std, rtcluster = 6)
```

<code>plotstdsda</code>	<i>Plot the std mass from GlobalStd algorithm in structure directed analysis(SDA) groups</i>
-------------------------	--

Description

Plot the std mass from GlobalStd algorithm in structure directed analysis(SDA) groups

Usage

```
plotstdsda(list, index = NULL, ...)
```

Arguments

`list` a list from `getsda` function
`index` index for PMD value
`...` other parameters for plot function

See Also

[getstd](#), [globalstd](#), [plotstd](#), [plotpaired](#), [plotstdrt](#)

Examples

```
data(spmein vivo)
re <- globalstd(spmein vivo, sda=TRUE)
plotstdsda(re)
```

pmdanno	<i>Perform MS/MS pmd annotation for mgf file</i>
---------	--

Description

Perform MS/MS pmd annotation for mgf file

Usage

```
pmdanno(file, db = NULL, ppm = 10, prems = 1.1, pmdc = 0.6, scutoff = 0.1)
```

Arguments

file	mgf file generated from MS/MS data
db	database could be list object from 'getms2pmd'
ppm	mass accuracy, default 10
prems	precursor mass range, default 1.1 to include M+H or M-H
pmdc	pmd length percentage cutoff for annotation. 0.6(default) means 60 percentage of the pmds in your sample could be found in certain compound pmd database
scutoff	relative intensity cutoff for input spectra for pmd analysis, default 0.1

Value

list with MSMS annotation results

runPMD	<i>Shiny application for PMD analysis</i>
--------	---

Description

Shiny application for PMD analysis

Usage

```
runPMD()
```

runPMDnet	<i>Shiny application for PMD network analysis</i>
-----------	---

Description

Shiny application for PMD network analysis

Usage

```
runPMDnet()
```

sda	<i>A dataset containing common Paired mass distances of substructure, ions replacements, and reaction</i>
-----	---

Description

A dataset containing common Paired mass distances of substructure, ions replacements, and reaction

Usage

```
data(sda)
```

Format

A data frame with 94 rows and 4 variables:

PMD Paired mass distances

origin potential sources

Ref. references

mode positive, negative or both mode to find corresponding PMDs

spmein vivo	<i>A peaks list dataset containing 9 samples from 3 fish with triplicates samples for each fish from LC-MS.</i>
-------------	---

Description

A peaks list dataset containing 9 samples from 3 fish with triplicates samples for each fish from LC-MS.

Usage

```
data(spmein vivo)
```

Format

A list with 4 variables from 1459 LC-MS peaks:

mz mass to charge ratios

rt retention time

data intensity matrix

group group information

Index

* datasets

hmdb, 14
keggrall, 15
omics, 15
sda, 22
spmein vivo, 22

sda, 22
spmein vivo, 22

getcda, 2
getchain, 3
getcluster, 4
getcorcluster, 5
getms2pmd, 5
getmpmd, 6
getpaired, 4, 6, 8, 10–12, 14, 17, 18
getpmd, 7, 10
getposneg, 8
getrda, 3, 8, 9, 10
getreact, 10
getsda, 3, 7–10, 11, 12, 14
getstd, 4, 7, 8, 10, 11, 12, 14, 18–20
gettarget, 13
globalstd, 13, 17–20

hmdb, 14

keggrall, 15

omics, 15

pcasf, 16
plotcn, 17
plotpaired, 7, 11, 17, 18, 20
plotrtg, 18
plotsda, 18
plotstd, 4, 12, 14, 18, 19, 20
plotstdrt, 14, 18, 19, 20
plotstdsda, 14, 20, 20
pmdanno, 21

runPMD, 21
runPMDnet, 21