Package 'pmd'

January 15, 2025

Type Package

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

Version 0.2.7

Description Paired mass distance (PMD) analysis proposed in Yu, Olkow-

icz and Pawliszyn (2018) <doi:10.1016/j.aca.2018.10.062> and PMD based reactomics analysis proposed in Yu and Petrick (2020) <doi:10.1038/s42004-020-00403-z> 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 VignetteBuilder knitr Depends R (>= 3.5.0) Imports RColorBrewer, stats, graphics, utils, igraph, enviGCMS RoxygenNote 7.3.2 NeedsCompilation no Author Miao YU [aut, cre] (<https://orcid.org/0000-0002-2804-6014>) Maintainer Miao YU <yufreecas@gmail.com> Repository CRAN Date/Publication 2025-01-15 19:40:05 UTC

Contents

getcda	2
getchain	3
getcluster	4
getcorcluster	5
getms2pmd	6
getmspmd	6
getpaired	7
getpmd	8
getpmddf	8
getposneg	9
getrda	10
getreact	11
getsda	12
getstd	13
gettarget	14
globalstd	14
hmdb	16
keggrall	16
MaConDa	17
omics	17
pcasf	18
ploten	19
plotpaired	19
plotrtg	20
<u>r</u>	20
F	21
plotstdrt	21
plotstdsda	22
	23
	23
	23
spmeinvivo	24
	~-
	25

Index

```
getcda
```

Perform correlation directed analysis for peaks list.

Description

Perform correlation directed analysis for peaks list.

Usage

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

getchain

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(spmeinvivo)
cluster <- getcorcluster(spmeinvivo)
cbp <- enviGCMS::getfilter(cluster,rowindex = cluster$stdmassindex2)
cda <- getcda(cbp)</pre>
```

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 0.6
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(spmeinvivo)
# check metabolites of C18H39N0
pmd <- getchain(spmeinvivo,diff = c(2.02,14.02,15.99),mass = 286.3101)
# remove the retention time for mass only data
spmeinvivo$rt <- NULL
pmd <- getchain(spmeinvivo,diff = c(2.02,14.02,15.99),mass = 286.3101)</pre>
```

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

getcorcluster

See Also

getpaired,getstd,plotstd

Examples

```
data(spmeinvivo)
re <- getpaired(spmeinvivo)
re <- getstd(re)
cluster <- getcluster(re)</pre>
```

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

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(spmeinvivo)
cluster <- getcorcluster(spmeinvivo)</pre>
```

getms2pmd

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
Scenspina	

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,
   mdrange = NULL
)
```

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
mdrange	mass defect range to ignore. Default NULL and $c(0.25,0.9)$ to retain the possible reaction related paired mass

Value

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

See Also

getstd,getsda,plotpaired

Examples

```
data(spmeinvivo)
pmd <- getpaired(spmeinvivo)</pre>
```

getpmd

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 or a vector of pmds
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 or pmds.

See Also

getpaired,getstd,getsda,getrda

Examples

```
data(spmeinvivo)
pmd <- getpmd(spmeinvivo,pmd=15.99)</pre>
```

getpmddf

Get pmd details for specific reaction after the removal of isotopouge.

Description

Get pmd details for specific reaction after the removal of isotopouge.

Usage

```
getpmddf(mz, group = NULL, pmd = NULL, digits = 2, mdrange = c(0.25, 0.9))
```

getposneg

Arguments

mz	a vector of mass to charge ratio.
group	mass to charge ratio group from either retention time or mass spectrometry imaging segmentation.
pmd	a specific paired mass distance or a vector of pmds
digits	mass or mass to charge ratio accuracy for pmd, default 2.
mdrange	mass defect range to ignore. Default $c(0.25,0.9)$ to retain the possible reaction related paired mass.

Value

dataframe with paired peaks for specific pmd or pmds. When group is provided, a column named net will be generated to show if certain pmd will be local(within the same group) or global(across the groups)

See Also

getpaired,getstd,getsda,getrda

Examples

```
data(spmeinvivo)
pmddf <- getpmddf(spmeinvivo$mz,pmd=15.99)</pre>
```

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

Description

Perform structure/reaction directed analysis for mass only.

Usage

```
getrda(
  mz,
  pmd = NULL,
  freqcutoff = 10,
  digits = 3,
  top = 20,
  formula = NULL,
  mdrange = c(0.25, 0.9),
  verbose = FALSE
)
```

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 automated
pmd	a specific paired mass distance or a vector of pmds, default NULL
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
mdrange	mass defect range to ignore. Default $c(0.25,0.9)$ to retain the possible reaction related paired mass
verbose	logic, if TURE, return will be llist with paired mass distances table. Default FALSE.

Value

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

See Also

getsda

getreact

Examples

```
data(spmeinvivo)
pmd <- getpaired(spmeinvivo)
std <- getstd(pmd)
sda <- getrda(spmeinvivo$mz[std$stdmassindex])
sda <- getrda(spmeinvivo$mz, pmd = c(2.016,15.995,18.011,14.016))</pre>
```

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,
  cvcutoff = 30,
  outlier = FALSE,
  method = "static",
  ...
)
```

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
cvcutoff	ratio or intensity cv cutoff for quantitative paired peaks, default 30
outlier	logical, if true, outlier of ratio will be removed, default False.
method	quantification method can be 'static' or 'dynamic'. See details.
	other parameters for getpmd

Details

PMD based reaction quantification methods have two options: 'static' will only consider the stable mass pairs across samples and such reactions will be limited by the enzyme or other factors than substrates. 'dynamic' will consider the unstable paired masses by normalization the relatively unstable peak with stable peak between paired masses and such reactions will be limited by one or both peaks in the paired masses.

Value

list with quantitative paired peaks.

See Also

getpaired,getstd,getsda,getrda,getpmd,

Examples

data(spmeinvivo)
pmd <- getreact(spmeinvivo,pmd=15.99)</pre>

getsda

Perform structure/reaction directed analysis for peaks list.

Description

Perform structure/reaction directed analysis for peaks list.

Usage

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

Arguments

list	a list with mzrt profile
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
freqcutoff	pmd frequency cutoff for structures or reactions, default NULL. This cutoff will be found by PMD network analysis when it is NULL.

12

getstd

Value

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

See Also

getpaired,getstd,plotpaired

Examples

```
data(spmeinvivo)
pmd <- getpaired(spmeinvivo)
std <- getstd(pmd)
sda <- getsda(std)</pre>
```

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(spmeinvivo)
pmd <- getpaired(spmeinvivo)
std <- getstd(pmd)</pre>
```

gettarget

Get multiple injections index for selected retention time

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(spmeinvivo)
pmd <- getpaired(spmeinvivo)
std <- getstd(pmd)
index <- gettarget(std$rt[std$stdmassindex])
table(index)</pre>
```

```
globalstd
```

GlobalStd algorithm with structure/reaction directed analysis

Description

GlobalStd algorithm with structure/reaction directed analysis

14

globalstd

Usage

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

Arguments

list	a peaks list with mass to charge, retention time and intensity data	
rtcutoff	cutoff of the distances in cluster, 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.	
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	
freqcutoff	pmd frequency cutoff for structures or reactions, default NULL. This cutoff will be found by PMD network analysis when it is NULL.	
mdrange	mass defect range to ignore. Default NULL and $c(0.25,0.9)$ to retain the possible reaction related paired mass	
sda	logical, option to perform structure/reaction directed analysis, default FALSE.	

Value

list with GlobalStd algorithm processed data.

See Also

getpaired,getstd,getsda,plotstd,plotstdsda,plotstdrt

Examples

```
data(spmeinvivo)
re <- globalstd(spmeinvivo)
```

hmdb

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

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

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 realted pmd and atoms numbers of C, H, O, N, P, S

ID KEGG reaction ID

pmd pmd with three digits

MaConDa

Description

mass spectrometry contaminants database for PMD check

Usage

data(MaConDa)

Format

A data frame from doi:10.1093/bioinformatics/bts527 with 308 rows and 5 variables:

id MaConDa ID

name contaminants

formula contaminants fomula

exact_mass exact mass of contaminants

type_of_contaminant type of contaminant

omics	A dataframe containing multiple reaction database ID and their re-
	lated accurate mass pmd and related reactions

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 realted 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

х	Matrix with sample in column and features in row
У	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)</pre>
```

18

plotcn

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('C6H1206','Glucose',c(2.016,14.016,15.995))

plotpaired

Plot the mass pairs and high frequency mass distances

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(spmeinvivo)
pmd <- getpaired(spmeinvivo)
plotpaired(pmd)</pre>

plotrtg

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(spmeinvivo)
pmd <- getpaired(spmeinvivo)
plotrtg(pmd)</pre>

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

plotstd

Examples

```
data(spmeinvivo)
re <- getpmd(spmeinvivo,pmd=78.9)
plotsda(re)</pre>
```

```
plotstd
```

Plot the std mass from GlobalStd algorithm

Description

Plot the std mass from GlobalStd algorithm

Usage

plotstd(list)

Arguments

list a list from getstd function

See Also

getstd, globalstd

Examples

```
data(spmeinvivo)
pmd <- getpaired(spmeinvivo)
std <- getstd(pmd)
plotstd(std)</pre>
```

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

Description

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

Usage

plotstdrt(list, rtcluster, ...)

plotstdsda

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(spmeinvivo)
pmd <- getpaired(spmeinvivo)
std <- getstd(pmd)
plotstdrt(std,rtcluster = 6)</pre>
```

plotstdsda	Plot the std mass from GlobalStd algorithm in structure directed anal-
procedusua	T for the sta mass from Giobaista argorithm in structure arrected anar-
	vsis(SDA) groups
	Join (2211) 8.0 mps

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(spmeinvivo)
re <- globalstd(spmeinvivo, sda=TRUE)
plotstdsda(re)</pre>
```

runPMD

Description

Shiny application for PMD analysis

Usage

runPMD()

runPMDnet

Shiny application for PMD network analysis

Description

Shiny application for PMD network analysis

Usage

runPMDnet()

sda

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

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

spmeinvivo

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

Description

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

Usage

data(spmeinvivo)

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, 16 keggrall, 16 MaConDa, 17 omics, 17 sda, 23 spmeinvivo, 24 getcda, 2getchain, 3 getcluster, 4 getcorcluster, 5 getms2pmd, 6 getmspmd, 6 getpaired, 5, 7, 8, 9, 12, 13, 15, 19, 20 getpmd, 8, 12 getpmddf, 8 getposneg, 9 getrda, 3, 8, 9, 10, 12 getreact, 11 getsda, 3, 7–10, 12, 12, 13, 15 getstd, 5, 7-9, 12, 13, 13, 15, 20-22 gettarget, 14globalstd, 14, 19-22 hmdb, 16 keggrall, 16 MaConDa, 17 omics, 17 pcasf, 18 plotcn, 19 plotpaired, 7, 13, 19, 20, 22 plotrtg, 20 plotsda, 20 plotstd, 5, 13, 15, 20, 21, 22 plotstdrt, 15, 20, 21, 22 plotstdsda, 15, 22, 22

runPMD, 23 runPMDnet, 23 sda, 23 spmeinvivo, 24