

Package: RAMClustR (via r-universe)

October 27, 2024

Type Package

Title Mass Spectrometry Metabolomics Feature Clustering and Interpretation

Version 1.3.0

Date 2023-05-17

Imports dynamicTreeCut, fastcluster, preprocessCore, e1071, gplots, pcaMethods, utils, ggplot2, InterpretMSSpectrum

Depends R (>= 3.5.0)

Suggests knitr, rmarkdown, xcms, testthat, BiocManager, stringi, curl

License MIT

Description A feature clustering algorithm for non-targeted mass spectrometric metabolomics data. This method is compatible with gas and liquid chromatography coupled mass spectrometry, including indiscriminant tandem mass spectrometry <DOI:10.1021/ac501530d> data.

URL <https://github.com/cbroeckl/RAMClustR>

Encoding UTF-8

biocViews MassSpectrometry, Metabolomics

RoxygenNote 7.2.3

VignetteBuilder knitr

Repository <https://cbroeckl.r-universe.dev>

RemoteUrl <https://github.com/cbroeckl/ramclustr>

RemoteRef HEAD

RemoteSha 5f0adcb2af21cda10217be50895bb540343e2de0

Contents

add_params	3
checks	3
check_arguments_filter.blanks	4
check_arguments_filter.cv	4

check_arguments_replace.na	5
compute_do.sets	5
compute_SpecAbundAve	6
compute_wt_mean	6
create_ramclustObj	7
defineExperiment	8
define_samples	9
doFindmain	9
exportDataset	12
filter_blanks	13
filter_good_features	13
filter_signal	14
findfeature	14
findmass	15
find_good_features	16
getData	17
get_ExpDes	18
get_instrument_platform	18
mean_signal_intensities	19
mergeRCobjects	19
normalized_data_batch_qc	21
normalized_data_tic	22
order_datasets	22
ramclustR	23
rc.calibrate.ri	26
rc.cmpd.filter.blanks	27
rc.cmpd.filter.cv	29
rc.cmpd.replace.na	30
rc.expand.sample.names	31
rc.export.msp.rc	32
rc.feature.filter.blanks	32
rc.feature.filter.cv	34
rc.feature.normalize.batch.qc	35
rc.feature.normalize.qc	36
rc.feature.normalize.quantile	37
rc.feature.normalize.tic	37
rc.feature.replace.na	38
rc.get.csv.data	39
rc.get.df.data	41
rc.get.xcms.data	43
rc.merge.split.clusters	44
rc.qc	45
rc.ramclustr	47
rc.remove.qc	48
rc.restore.qc.samples	49
RCQC	50
remove_blanks	51
replace_na	52

<i>add_params</i>	3
write.gcei.mat	52
write.methods	53
write.msp	54
write_csv	54

Index **56**

add_params	<i>add_params</i>
------------	-------------------

Description

add rc.feature.replace.na params in ramclustObj

Usage

```
add_params(ramclustObj, params, param_name)
```

Arguments

ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
params	vector containing parameters to add
param_name	name of the parameter/step

Value

ramclustR object with rc.feature.replace.na params added.

checks	<i>checks</i>
--------	---------------

Description

check if MS data contains mz and rt, and if MSMS data is present feature names and sample names are identical

Usage

```
checks(
  ms1_featureDefinitions = NULL,
  ms1_featureValues = NULL,
  ms2_featureValues = NULL,
  feature_names = NULL
)
```

Arguments

ms1_featureDefinitions	dataframe with metadata with columns: mz, rt, feature names containing MS data
ms1_featureValues	dataframe with rownames = sample names, colnames = feature names containing MS data
ms2_featureValues	dataframe with rownames = sample names, colnames = feature names containing MSMS data
feature_names	feature names extracted from the data

check_arguments_filter.blanks
check_arguments_filter.blanks

Description

check provided arguments

Usage

check_arguments_filter.blanks(ramclustObj, sn)

Arguments

ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
sn	numeric defines the ratio for 'signal'. i.e. sn = 3 indicates that signal intensity must be 3 fold higher in sample than in blanks, on average, to be retained.

check_arguments_filter.cv
check_arguments_filter.cv

Description

check provided arguments

Usage

check_arguments_filter.cv(ramclustObj, qc.tag)

Arguments

ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
qc.tag	character vector of length one or two. If length is two, enter search string and factor name in \$phenoData slot (i.e. c("QC", "sample.type"). If length one (i.e. "QC"), will search for this string in the 'sample.names' slot by default.

```
check_arguments_replace.na
      check_arguments_replace.na
```

Description

check provided arguments

Usage

```
check_arguments_replace.na(
  ramclustObj,
  replace.int,
  replace.noise,
  replace.zero
)
```

Arguments

ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
replace.int	default = 0.1. proportion of minimum feature value to replace NA (or zero) values with
replace.noise	default = 0.1. proportion of replace.int value by which noise is added via 'jitter'
replace.zero	logical if TRUE, any zero values are replaced with noise as if they were NA values

```
compute_do.sets      compute_do.sets
```

Description

compute data frame to use in ramclustObj

Usage

```
compute_do.sets(ramclustObj)
```

Arguments

ramclustObj ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)

Value

vector which is used to select data frame to use in ramclustObj

compute_SpecAbundAve *compute_SpecAbundAve*

Description

further aggregate by sample names for 'SpecAbundAve' dataset

Usage

```
compute_SpecAbundAve(ramclustObj = NULL)
```

Arguments

ramclustObj ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)

Value

ramclustR object with aggregate by sample names for 'SpecAbundAve' dataset

compute_wt_mean *compute_wt_mean*

Description

compute weighted.mean intensity of feature in ms/msms level data

Usage

```
compute_wt_mean(data, global.min, fmz, ensure.no.na)
```

Arguments

data	feature in ms/msms level data
global.min	minimum intensity in ms/msms level data
fmz	feature retention time
ensure.no.na	logical: if TRUE, any 'NA' values in msint and/or msmsint are replaced with numerical values based on 10 percent of feature min plus noise. Used to ensure that spectra are not written with NA values.

Value

weighted.mean intensity of feature in ms/msms level data

create_ramclustObj *create_ramclustObj*

Description

create ramclustr Object

Usage

```
create_ramclustObj(
  ExpDes = NULL,
  input_history = NULL,
  MSdata = NULL,
  MSMSdata = NULL,
  frt = NULL,
  fmz = NULL,
  st = NULL,
  phenoData = NULL,
  feature_names = NULL,
  sample_names = NULL,
  xcmsOrd = NULL,
  ensure.no.na = TRUE
)
```

Arguments

ExpDes	either an R object created by R ExpDes object: data used for record keeping and labelling msp spectral output
input_history	input history
MSdata	dataframe containing MS Data
MSMSdata	dataframe containing MSMS Data
frt	feature retention time, in whatever units were fed in
fmz	feature retention time
st	numeric: sigma t - time similarity decay value
phenoData	dataframe containing phenoData
feature_names	feature names extracted from the data
sample_names	sample names extracted from the data
xcmsOrd	original xcms order of features, for back-referencing when necessary
ensure.no.na	logical: if TRUE, any 'NA' values in msint and/or msmsint are replaced with numerical values based on 10 percent of feature min plus noise. Used to ensure that spectra are not written with NA values.

Value

an ramclustR object. this object is formatted as an hclust object with additional slots for holding feature and compound data.

defineExperiment *defineExperiment*

Description

Create an Experimental Design R object for record-keeping and msp output

Usage

```
defineExperiment(csv = FALSE, force.skip = FALSE)
```

Arguments

csv	logical or filepath. If csv = TRUE , csv template called "ExpDes.csv" will be written to your working directory. you will fill this in manually, ensuring that when you save you retain csv format. ramclustR will then read this file in and and format appropriately. If csv = FALSE, a pop up window will appear (in windows, at least) asking for input. If a character string with full path (and file name) to a csv file is given, this will allow you to read in a previously edited csv file.
force.skip	logical. If TRUE, ramclustR creates a pseudo-filled ExpDes object to enable testing of functionality. Not recommended for real data, as your exported spectra will be improperly labelled.

Value

an Exp Des R object which will be used for record keeping and writing spectra data.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. *Anal Chem*. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

define_samples	<i>define_samples</i>
----------------	-----------------------

Description

define samples in each set

Usage

```
define_samples(ramclustObj, tag, return.logical = FALSE)
```

Arguments

ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
tag	character vector of length one or two. If length is two, enter search string and factor name in \$phenoData slot (i.e. c("QC", "sample.type"). If length one (i.e. "QC"), will search for this string in the 'sample.names' slot by default.
logical	optionally convert numeric vector with length equal to the number of matched samples to a logical vector of length equal to number of samples, with TRUE representing matching samples.

Value

samples found using the tag

doFindmain	<i>doFindmain</i>
------------	-------------------

Description

Cluster annotation function: inference of 'M' - molecular weight of the compound giving rise to each spectrum - using the InterpretMSSpectrum::findMain function

Usage

```
doFindmain(  
  ramclustObj = NULL,  
  compd = NULL,  
  mode = "positive",  
  mzabs.error = 0.005,  
  ppm.error = 10,  
  ads = NULL,  
  nls = NULL,  
  scoring = "auto",  
  plot.findmain = TRUE,
```

```

writeMat = TRUE,
writeMS = TRUE,
use.z = TRUE
)

```

Arguments

ramclustObj	ramclustR object to annotate.
compd	integer: vector defining compound numbers to annotated. if NULL (default), all compounds
mode	character: "positive" or "negative"
mzabs.error	numeric: absolute mass deviation allowed, default = 0.01
ppm.error	numeric: ppm mass error <code>_added_</code> to <code>mzabs.error</code> , default = 10
ads	character: vector of allowed adducts, i.e. <code>c("[M+H]+")</code> . if NULL, default positive mode values of H+, Na+, K+, and NH ₄ +, as monomer, dimer, and trimer, are assigned. Negative mode include "[M-H]-", "[M+Na-2H]-", "[M+K-2H]-", "[M+CH ₂ O ₂ -H]-" as monomer, dimer, and trimer.
nls	character: vector of allowed neutral losses, i.e. <code>c("[M+H-H₂O]+")</code> . if NULL, an extensive list derived from CAMERA's will be used.
scoring	character: one of 'imss', 'ramclustr', or 'auto'. default = 'auto'. see details.
plot.findmain	logical: should pdf plots be generated for evaluation? default = TRUE. PDF saved to <code>working.directory/spectra</code>
writeMat	logical: should individual .mat files (for MSFinder) be generated in a 'mat' subdirectory in the 'spectra' folder? default = TRUE.
writeMS	logical: should individual .ms files (for Sirius) be generated in a 'ms' subdirectory in the 'spectra' folder? default = TRUE. Note that no import functions are yet written for Sirius output.
use.z	logical: if you have previously run the 'assign.z' function from ramclustR, there will be a slot reflecting the feature mass after accounting for charge (fm) - if TRUE this is used instead of feature m/z (fmz) in interpreting MS data and exporting spectra for annotation.

Details

a partially annotated ramclustR object. base structure is that of a standard R hierarchical clustering output, with additional slots described in ramclustR documentation (?ramclustR). New slots added after using the interpretMSSpectrum functionality include those described below.

Value

\$M: The inferred molecular weight of the compound giving rise to the each spectrum

\$M.ppm: The ppm error of all the MS signals annotated, high error values should be considered 'red flags'.

\$M.ann: The annotated spectrum supporting the interpretation of M

`$use.findmain`: Logical vector indicating whether findmain scoring (TRUE) or ramclustR scoring (FALSE) was used to support inference of M. By default, findmain scoring is used. When ramclustR scoring differs from findmain scoring, the scoring metric which predicts higher M is selected.

`$M.ramclustr`: M selected using ramclustR scoring

`$M.ppm.ramclustr`: ppm error of M selected using ramclustR scoring. Used to resolve conflicts between ramclustR and findmain M assignment when scoring = auto.

`$M.ann.ramclustr`: annotated spectrum supporting M using ramclustR scoring

`$M.nann.ramclustr`: number of masses annotated using ramclustR scoring. Used to resolve conflicts between ramclustR and findmain M assignment when scoring = auto.

`$M.space.ramclustr`: the 'space' of scores between the best and second best ramclustR scores. Calculated as a ratio. Used to resolve conflicts between ramclustR and findmain M assignment when scoring = auto.

`$M.findmain`: M selected using findmain scoring

`$M.ppm.findmain`: ppm error of M selected using findmain scoring. Used to resolve conflicts between ramclustR and findmain M assignment when scoring = auto.

`$M.ann.findmain`: annotated spectrum supporting M using findmain scoring

`$M.nann.findmain`: number of masses annotated using findmain scoring. Used to resolve conflicts between ramclustR and findmain M assignment when scoring = auto.

`$M.space.findmain`: the 'space' of scores between the best and second best findmain scores. Calculated as a ratio. Used to resolve conflicts between ramclustR and findmain M assignment when scoring = auto.

Author(s)

Corey Broeckling

References

Jaeger C, ... Lisek J. Compound annotation in liquid chromatography/high-resolution mass spectrometry based metabolomics: robust adduct ion determination as a prerequisite to structure prediction in electrospray ionization mass spectra. *Rapid Commun Mass Spectrom*. 2017 Aug 15;31(15):1261-1266. doi: 10.1002/rcm.7905. PubMed PMID: 28499062.

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. *Anal Chem*. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

exportDataset	<i>exportDataset</i>
---------------	----------------------

Description

export one of 'SpecAbund', 'SpecAbundAve', 'MSdata' or 'MSMSdata' from an RC object to csv

Usage

```
exportDataset(  
  ramclustObj = NULL,  
  which.data = "SpecAbund",  
  label.by = "ann",  
  appendFactors = TRUE  
)
```

Arguments

ramclustObj	ramclustR object to export from
which.data	name of dataset to export. SpecAbund, SpecAbundAve, MSdata, or MSMSdata
label.by	either 'ann' or 'cmpd', generally. name of ramclustObj slot used as csv header for each column (compound)
appendFactors	logical. If TRUE (default) the factor data frame is appended to the left side of the dataset.

Details

Useful for exporting the processed signal intensity matrix to csv for analysis elsewhere.

Value

nothing is returned. file exported as csv to 'datasets/*.csv'

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

filter_blanks	<i>filter_blanks</i>
---------------	----------------------

Description

filter blanks

Usage

```
filter_blanks(ramclustObj, keep, d1)
```

Arguments

ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
keep	union of which signal is at least 3x larger, output of filter_signal()
d1	MS Data

Value

ramclustObj object with feature.filter.blanks

filter_good_features	<i>filter_good_features</i>
----------------------	-----------------------------

Description

filter to keep only 'good' features

Usage

```
filter_good_features(ramclustObj, keep)
```

Arguments

ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
keep	features to keep. output of find_good_features().

Value

ramclustR object filtered to keep only 'good' features

filter_signal	<i>filter_signal</i>
---------------	----------------------

Description

filter signal

Usage

```
filter_signal(ms.qc.mean, ms.blank.mean, sn)
```

Arguments

ms.qc.mean	ms qc mean signal intensities
ms.blank.mean	ms blank mean signal intensities
sn	numeric defines the ratio for 'signal'. i.e. sn = 3 indicates that signal intensity must be 3 fold higher in sample than in blanks, on average, to be retained.

Value

union of which signal is at least 3x larger

findfeature	<i>findfeature</i>
-------------	--------------------

Description

see if any features match a given mass, and whether they are plausibly M0

Usage

```
findfeature(  
  ramclustObj = NULL,  
  mz = NULL,  
  mztol = 0.02,  
  rt = NULL,  
  rttol = 2,  
  iso.rttol = 2,  
  zmax = 6,  
  m.check = TRUE  
)
```

Arguments

ramclustObj	R object: the ramclustR object to explore
mz	numeric: mz value to search for
mztol	numeric: absolute mass tolerance around mz
rt	numeric: optional rt value to search for (generally in seconds, though use whatever units your data is in)
rttol	numeric: absolute retention time tolerance around rt.
iso.rttol	numeric: when examining isotope patterns, feature retention time tolerance around features matching mz +/- mztol
zmax	integer: maximum charge state to consider. default is 6.
m.check	logical: check whether the matching masses are plausibly M0. That is, we look for ions 1 proton mass (from charge state 1:zmax) below the target m/z at the same time that have intensities consistent with target ion being a non-M0 isotope.

Details

a convenience function to perform a targeted search of all features for a mass of interest. Also performs a crude plausibility check as to whether the matched feature could be M0, based on the assumption of approximately 1 carbon per 17 m/z units and natural isotopic abundance of 1.1

Value

returns a table to the console listing masses which match, their retention time and intensity, and whether it appears to be plausible as M0

Author(s)

Corey Broeckling

findmass

findmass

Description

see if any features match a given mass, and whether they are plausibly M0

Usage

```
findmass(  
  ramclustObj = NULL,  
  mz = NULL,  
  mztol = 0.02,  
  rttol = 2,  
  zmax = 6,  
  m.check = TRUE  
)
```

Arguments

ramclustObj	R object: the ramclustR object to explore
mz	numeric: mz value to search for
mztol	numeric: absolute mass tolerance around mz
rttol	numeric: when examining isotope patterns, feature retention time tolerance around features matching mz +- mztol
zmax	integer: maximum charge state to consider. default is 6.
m.check	logical: check whether the matching masses are plausibly M0. That is, we look for ions 1 proton mass (from charge state 1:zmax) below the target m/z at the same time that have intensities consistent with target ion being a non-M0 isotope.

Details

a convenience function to perform a targeted search of all feaures for a mass of interest. Also performs a crude plausibility check as to whether the matched feature could be M0, based on the assumption of approximately 1 carbon per 17 m/z units and natural isotopic abundance of 1.1

Value

returns a table to the console listing masses which match, their retention time and intensity, and whether it appears to be plausible as M0

Author(s)

Corey Broeckling

find_good_features *find_good_features*

Description

find 'good' features, acceptable CV at either MS or MSMS level results in keeping

Usage

```
find_good_features(ramclustObj, do.sets, max.cv, qc)
```

Arguments

ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
do.sets	select data frame to use.
max.cv	numeric maximum allowable cv for any feature. default = 0.5
qc	QC samples found by define_samples

Value

ramclustR object
features to keep

getData

getData

Description

retrieve and parse sample names, retrieve metabolite data. returns as list of two data frames

Usage

```
getData(  
  ramclustObj = NULL,  
  which.data = "SpecAbund",  
  delim = "-",  
  cmpdlabel = "cmpd",  
  filter = FALSE  
)
```

Arguments

ramclustObj	ramclustR object to retrieve data from
which.data	character; which dataset (SpecAbund or SpecAbundAve) to reference
delim	character; "-" by default - the delimiter for parsing sample names to factors
cmpdlabel	= "cmpd"; label the data with the annotation. can also be set to 'ann' for column names assigned as annotations.
filter	= TRUE; logical, if TRUE, checks for \$cmpd.use slot generated by rc.cmpd.cv.filter() function, and only gets acceptable compounds.

Details

convenience function for parsing sample names and returning a dataset.

Value

returns a list of length 3: \$design is the experimental sample factors after parsing by the delim, \$data is the dataset, \$full.data is merged \$des and \$data data.frames.

Author(s)

Corey Broeckling

`get_ExpDes` *get_ExpDes*

Description

get Experimental Design

Usage

`get_ExpDes(csv.in)`

Arguments

`csv.in` Experimental Design read from csv

Value

list containing design and instrument

`get_instrument_platform`
get_instrument_platform

Description

get instrument platform

Usage

`get_instrument_platform(design)`

Arguments

`design` data frame containing Experimental Design

Value

instrument platform

```
mean_signal_intensities  
    mean_signal_intensities
```

Description

calculate MS mean signal intensities

Usage

```
mean_signal_intensities(data, sample)
```

Arguments

data	MS/MSMS data
sample	sample found using the tag, output of define_samples()

Value

mean signal intensities

```
mergeRCobjects    mergeRCobjects
```

Description

merge two ramclustR objects

Usage

```
mergeRCobjects(  
  ramclustObj.1 = NULL,  
  ramclustObj.2 = NULL,  
  mztol = 0.02,  
  rttol = 30,  
  course.rt.adj = NULL,  
  mzwt = 2,  
  rtw = 1,  
  intwt = 3  
)
```

Arguments

ramclustObj.1	ramclustR object 1: this object will be the base for the new object. That is all the features from ramclustObj.1 will be retained.
ramclustObj.2	ramclustR object 2: this object will mapped and appended to ramclustObj.1. That is only features which appear consistent with those from ramclustObj.1 will be retained.
mztol	numeric: absolute mass tolerance around mz
rttol	numeric: feature retention time tolerance. Value set by this option will be used during the initial anchor mapping phase. Two times the standard error of the rt loess correction will be used for the full mapping.
course.rt.adj	numeric: default = NULL. optional approximate retention time shift between ramclustObj.1 and ramclustObj.2. i.e if the retention time of ramclustObj.1 is on average 15 seconds longer than that of ramclustObj.2, enter '15'. if 1 is less than 2, enter a negative number. This is applied before mapping to enable a smaller 'rttol' value to be used.
mzwt	numeric: when mapping features, weighting value used for similarities between feature mass values (see rtwt, intwt)
rtwt	numeric: when mapping features, weighting value used for similarities between feature retention time values (see mzwt, intwt)
intwt	numeric: when mapping features, weighting value used for similarities between ranked signal intensity values (see rtwt, mzwt)

Details

Two ramclustR objects are merged with this function, mapping features between them. The first (ramclustObj.1) object use used as the template - all data in it is retained. ramclustObj.2 is mapped to ramclustObj.1 feature by feature - only mapped features are retained. A new ramclustObj is returned, with a new SpecAbund dataset with the same column number as the ramclustObj.1\$SpecAbund set.

Value

returns a ramclustR object. All values from ramclustObj.1 are retained. SpecAbund dataset from ramclustObj.1 is moved to RC\$SpecAbund.1, where RC is the new ramclustObj.

Author(s)

Corey Broeckling

```
normalized_data_batch_qc  
    normalized_data_batch_qc
```

Description

normalize data using batch.qc

Usage

```
normalized_data_batch_qc(  
  data = NULL,  
  batch = NULL,  
  order = NULL,  
  qc = NULL,  
  qc.inj.range = 20  
)
```

Arguments

data	feature in ms/msms level data
batch	integer vector with length equal to number of injections in xset or csv file or dataframe
order	integer vector with length equal to number of injections in xset or csv file or dataframe
qc	logical vector with length equal to number of injections in xset or csv file or dataframe
qc.inj.range	integer: how many injections around each injection are to be scanned for presence of QC samples when using batch.qc normalization? A good rule of thumb is between 1 and 3 times the typical injection span between QC injections. i.e. if you inject QC ever 7 samples, set this to between 7 and 21. smaller values provide more local precision but make normalization sensitive to individual poor outliers (though these are first removed using the boxplot function outlier detection), while wider values provide less local precision in normalization but better stability to individual peak areas.

Value

normalized data.

normalized_data_tic *normalized_data_tic*

Description

normalize data using TIC

Usage

```
normalized_data_tic(ramclustObj = NULL)
```

Arguments

ramclustObj ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)

Value

ramclustR object with total extracted ion normalized data.

order_datasets *order_datasets*

Description

order the datasets first by batch and run order

Usage

```
order_datasets(order = NULL, batch = NULL, qc = NULL, data = NULL)
```

Arguments

order	integer vector with length equal to number of injections in xset or csv file or dataframe
batch	integer vector with length equal to number of injections in xset or csv file or dataframe
qc	logical vector with length equal to number of injections in xset or csv file or dataframe
data	feature in ms/msms level data

Value

ordered feature in ms/msms level data, order, batch, qc

ramclustR	<i>ramclustR</i>
-----------	------------------

Description

Main clustering function for grouping features based on their analytical behavior.

Usage

```
ramclustR(  
  xcmsObj = NULL,  
  ms = NULL,  
  pheno_csv = NULL,  
  idsms = NULL,  
  taglocation = "filepaths",  
  MStag = NULL,  
  idMSMStag = NULL,  
  featdelim = "_",  
  timepos = 2,  
  st = NULL,  
  sr = NULL,  
  maxt = NULL,  
  deepSplit = FALSE,  
  blocksize = 2000,  
  mult = 5,  
  hmax = NULL,  
  sampNameCol = 1,  
  collapse = TRUE,  
  usePheno = TRUE,  
  mspout = TRUE,  
  ExpDes = NULL,  
  normalize = "TIC",  
  qc.inj.range = 20,  
  order = NULL,  
  batch = NULL,  
  qc = NULL,  
  minModuleSize = 2,  
  linkage = "average",  
  mzdec = 3,  
  cor.method = "pearson",  
  rt.only.low.n = TRUE,  
  replace.zeros = TRUE  
)
```

Arguments

`xcmsObj` `xcmsObject`: containing grouped feature data for clustering by `ramclustR`

ms	filepath: optional csv input. Features as columns, rows as samples. Column header mz_rt
pheno_csv	filepath: optional csv input containing phenoData
idmsms	filepath: optional idMSMS / MSe csv data. same dim and names as ms required
taglocation	character: "filepaths" by default, "phenoData[,1]" is another option. refers to xcms slot
MStag	character: character string in 'taglocation' to designat MS / MSe files e.g. "01.cdf"
idMSMStag	character: character string in 'taglocation' to designat idMSMS / MSe files e.g. "02.cdf"
featdelim	character: how feature mz and rt are delimited in csv import column header e.g. "="
timepos	integer: which position in delimited column header represents the retention time (csv only)
st	numeric: sigma t - time similarity decay value
sr	numeric: sigma r - correlational similarity decay value
maxt	numeric: maximum time difference to calculate retention similarity for - all values beyond this are assigned similarity of zero
deepSplit	logical: controls how aggressively the HCA tree is cut - see ?cutreeDynamicTree
blocksize	integer: number of features (scans?) processed in one block =1000,
mult	numeric: internal value, can be used to influence processing speed/ram usage
hmax	numeric: precut the tree at this height, default 0.3 - see ?cutreeDynamicTree
sampNameCol	integer: which column from the csv file contains sample names?
collapse	logical: reduce feature intensities to spectrum intensities?
usePheno	logical: transfer phenotype data from XCMS object to SpecAbund dataset?
mspout	logical: write msp formatted spectra to file?
ExpDes	either an R object created by R ExpDes object: data used for record keeping and labelling msp spectral output
normalize	character: either "none", "TIC", "quantile", or "batch.qc" normalization of feature intensities. see batch.qc overview in details.
qc.inj.range	integer: how many injections around each injection are to be scanned for presence of QC samples when using batch.qc normalization? A good rule of thumb is between 1 and 3 times the typical injection span between QC injections. i.e. if you inject QC ever 7 samples, set this to between 7 and 21. smaller values provide more local precision but make normalization sensitive to individual poor outliers (though these are first removed using the boxplot function outlier detection), while wider values provide less local precision in normalization but better stability to individual peak areas.
order	integer vector with length equal to number of injections in xset or csv file
batch	integer vector with length equal to number of injections in xset or csv file
qc	logical vector with length equal to number of injections in xset or csv file.
minModuleSize	integer: how many features must be part of a cluster to be returned? default = 2

linkage	character: heirarchical clustering linkage method - see ?hclust
mzdec	integer: number of decimal places used in printing m/z values
cor.method	character: which correlational method used to calculate 'r' - see ?cor
rt.only.low.n	logical: default = TRUE At low injection numbers, correlational relationships of peak intensities may be unreliable. by default ramclustR will simply ignore the correlational r value and cluster on retention time alone. if you wish to use correlation with at n < 5, set this value to FALSE.
replace.zeros	logical: TRUE by default. NA, NaN, and Inf values are replaced with zero, and zero values are sometimes returned from peak peaking. When TRUE, zero values will be replaced with a small amount of noise, with noise level set based on the detected signal intensities for that feature.

Details

Main clustering function output - see citation for algorithm description or vignette('RAMClustR') for a walk through. batch.qc. normalization requires input of three vectors (1) batch (2) order (3) qc. This is a feature centric normalization approach which adjusts signal intensities first by comparing batch median intensity of each feature (one feature at a time) QC signal intensity to full dataset median to correct for systematic batch effects and then secondly to apply a local QC median vs global median sample correction to correct for run order effects.

Value

\$featclus: integer vector of cluster membership for each feature

\$firt: feature retention time, in whatever units were fed in (xcms uses seconds, by default)

\$fmz: feature retention time, reported in number of decimal points selected in ramclustR function

\$xcmsOrd: the original XCMS (or csv) feature order for cross referencing, if need be

\$clrt: cluster retention time

\$clrtsd: retention time standard deviation of all the features that comprise that cluster

\$nfeat: number of features in the cluster

\$nsing: number of 'singletons' - that is the number of features which clustered with no other feature

\$ExpDes: the experimental design object used when running ramclustR. List of two dataframes.

\$cmpd: compound name. C#### are assigned in order of output by dynamicTreeCut. Compound with the most features is classified as C0001...

\$ann: annotation. By default, annotation names are identical to 'cmpd' names. This slot is a placeholder for when annotations are provided

\$MSdata: the MSdataset provided by either xcms or csv input

\$MSMSdata: the (optional) MSe/idMSMS dataset provided be either xcms or csv input

\$SpecAbund: the cluster intensities after collapsing features to clusters

\$SpecAbundAve: the cluster intensities after averaging all samples with identical sample names

- 'spectra' directory is created in the working directory. In this directory a .msp is (optionally) created, which contains the spectra for all compounds in the dataset following clustering. if MSe/idMSMS data are provided, they are listed with the same compound name as the MS spectrum, with the collision energy provided in the ExpDes object provided to distinguish low from high CE spectra.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. *Anal Chem*. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

Examples

```
## Choose input file with feature column names `mz_rt` (expected by default).
## Column with sample name is expected to be first (by default).
## These can be adjusted with the `featdelim` and `sampNameCol` parameters.
wd <- getwd()
filename <- system.file("extdata", "peaks.csv", package = "RAMClustR", mustWork = TRUE)
print(filename)
head(data.frame(read.csv(filename)), c(6L, 5L))

## If the file contains features from MS1, assign those to the `ms` parameter.
## If the file contains features from MS2, assign those to the `idmsms` parameter.
## If you ran `xcms` for the feature detection, the assign the output to the `xcmsObj` parameter.
## In this example we use a MS1 feature table stored in a `csv` file.
setwd(tempdir())
ramclustobj <- ramclustR(ms = filename, st = 5, maxt = 1, blocksize = 1000)

## Investigate the deconvoluted features in the `spectra` folder in MSP format
## or inspect the `ramclustobj` for feature retention times, annotations etc.
print(ramclustobj$ann)
print(ramclustobj$nfeat)
print(ramclustobj$SpecAbund[, 1:6])
setwd(wd)
```

rc.calibrate.ri

rc.calibrate.ri

Description

extractor for xcms objects in preparation for clustering

Usage

```
rc.calibrate.ri(ramclustObj = NULL, calibrant.data = "", poly.order = 3)
```

Arguments

ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
calibrant.data	character vector defining the file path/name to a csv file containing columns including 'rt', and 'ri'. Alternatively, a data.frame with those column names (case sensitive)
poly.order	integer default = 3. polynomial order used to fit rt vs ri data, and calculate ri for all feature and metabolite rt values. poly.order should be appreciably smaller than the number of calibrant points.

Details

This function generates a new slot in the ramclustR object for retention index. Calibration is performed using a polynomial fit of order poly.order. It is the user's responsibility to ensure that the number and span of calibrant points is sufficient to calibrate the full range of feature and compound retention times. i.e. if the last calibration point is at 1000 seconds, but the last eluting peak is at 1300 seconds, the calibration will be very poor for the late eluting compound.

Value

ramclustR object with retention index assigned for features (\$fri) and compounds (\$cri).

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

rc.cmpd.filter.blanks *rc.cmpd.filter.blanks*

Description

used to remove compounds which are found at similar intensity in blank samples. Only applied after clustering. see also rc.feature.filter.blanks for filtering at the feature level (only done before clustering).

Usage

```
rc.cmpd.filter.blanks(  
  ramclustObj = NULL,  
  qc.tag = "QC",  
  blank.tag = "blank",  
  sn = 3,  
  remove.blanks = TRUE  
)
```

Arguments

ramclustObj	ramclustObj containing SpecAbund dataframe.
qc.tag	character vector of length one or two. If length is two, enter search string and factor name in \$phenoData slot (i.e. c("QC", "sample.type"). If length one (i.e. "QC"), will search for this string in the 'sample.names' slot by default.
blank.tag	see 'qc.tag' , but for blanks to use as background.
sn	numeric defines the ratio for 'signal'. i.e. sn = 3 indicates that signal intensity must be 3 fold higher in sample than in blanks, on average, to be retained.
remove.blanks	logical. TRUE by default. this removes any recognized blanks samples from the SpecAbund sets after they are used to filter contaminant compounds

Details

This function removes compounds which contain signal in QC samples comparable to blanks.

Value

ramclustR object with normalized data.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

rc.cmpd.filter.cv *rc.cmpd.filter.cv*

Description

extractor for xcms objects in preparation for clustering

Usage

```
rc.cmpd.filter.cv(ramclustObj = NULL, qc.tag = "QC", max.cv = 0.5)
```

Arguments

ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
qc.tag	character vector of length one or two. If length is two, enter search string and factor name in \$phenoData slot (i.e. c("QC", "sample.type"). If length one (i.e. "QC"), will search for this string in the 'sample.names' slot by default.
max.cv	numeric maximum allowable cv for any feature. default = 0.3

Details

This function offers normalization by total extracted ion signal. it is recommended to first run 'rc.feature.filter.blanks' to remove non-sample derived signal.

Value

ramclustR object with total extracted ion normalized data.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem.* 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

`rc.cmpd.replace.na` *rc.cmpd.replace.na*

Description

replaces any NA (and optionally zero) values with small signal (20

Usage

```
rc.cmpd.replace.na(  
  ramclustObj = NULL,  
  replace.int = 0.1,  
  replace.noise = 0.1,  
  replace.zero = TRUE  
)
```

Arguments

<code>ramclustObj</code>	ramclustObj containing SpecAbund dataset
<code>replace.int</code>	default = 0.2. proportion of minimum feature value to replace NA (or zero) values with
<code>replace.noise</code>	default = 0.2. proportion of <code>replace.int</code> value by which noise is added via 'jitter'
<code>replace.zero</code>	logical if TRUE, any zero values are replaced with noise as if they were NA values

Details

noise is added by finding for each feature the minimum detected value, multiplying that value by `replace.int`, then adding (`replace.int*replace.noise`) noise. `abs()` is used to ensure no negative values result.

Value

ramclustR object with NA and zero values removed.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

```
rc.expand.sample.names  
rc.expand.sample.names
```

Description

turn concatenated sample names into factors

Usage

```
rc.expand.sample.names(  
  ramclustObj = NULL,  
  delim = "-",  
  factor.names = TRUE,  
  quiet = FALSE  
)
```

Arguments

ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
delim	what delimiter should be used to separate names into factors? '-' by default
factor.names	logical or character vector. if TRUE, user will enter names one by one in console. If character vector (i.e. c("trt", "time")) names are assigned to table
quiet	logical . if TRUE, user will not be prompted to enter names one by one in console.

Details

THIS function only works on newer format ramclustObjects with a \$phenoData slot.

This function will split sample names by a delimiter, and enable users to name factors

Value

ramclustR object with normalized data.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

rc.export.msp.rc *rc.export.msp.rc*

Description

Cluster annotation function: inference of 'M' - molecular weight of the compound giving rise to each spectrum - using the InterpretMSSpectrum::findMain function

Usage

```
rc.export.msp.rc(ramclustObj = NULL, one.file = TRUE, mzdec = 1)
```

Arguments

ramclustObj	ramclustR object to annotate.
one.file	logical, should all msp spectra be written to one file? If false, each spectrum is an individual file.
mzdec	integer. Number of decimal points to export mass values with.

Details

exports files to a directory called 'spectra'. If one.file = FALSE, a new directory 'spectra/msp' is created to hold the individual msp files. if do.findman has been run, spectra are written as ms2 spectra, else as ms1.

Value

nothing, just exports files to the working directory

Author(s)

Corey Broeckling

rc.feature.filter.blanks
rc.feature.filter.blanks

Description

used to remove features which are found at similar intensity in blank samples

Usage

```
rc.feature.filter.blanks(  
  ramclustObj = NULL,  
  qc.tag = "QC",  
  blank.tag = "blank",  
  sn = 3,  
  remove.blanks = TRUE  
)
```

Arguments

ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
qc.tag	character vector of length one or two. If length is two, enter search string and factor name in \$phenoData slot (i.e. c("QC", "sample.type"). If length one (i.e. "QC"), will search for this string in the 'sample.names' slot by default.
blank.tag	see 'qc.tag' , but for blanks to use as background.
sn	numeric defines the ratio for 'signal'. i.e. sn = 3 indicates that signal intensity must be 3 fold higher in sample than in blanks, on average, to be retained.
remove.blanks	logical. TRUE by default. this removes any recognized blanks samples from the MSdata and MSMSdata sets after they are used to filter contaminant features.

Details

This function offers normalization by run order, batch number, and QC sample signal intensity.

Each input vector should be the same length, and equal to the number of samples in the \$MSdata set.

Input vector order is assumed to be the same as the sample order in the \$MSdata set.

Value

ramclustR object with normalized data.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

rc.feature.filter.cv *rc.feature.filter.cv*

Description

extractor for xcms objects in preparation for clustering

Usage

```
rc.feature.filter.cv(ramclustObj = NULL, qc.tag = "QC", max.cv = 0.5)
```

Arguments

ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
qc.tag	character vector of length one or two. If length is two, enter search string and factor name in \$phenoData slot (i.e. c("QC", "sample.type"). If length one (i.e. "QC"), will search for this string in the 'sample.names' slot by default.
max.cv	numeric maximum allowable cv for any feature. default = 0.5

Details

This function offers normalization by total extracted ion signal. it is recommended to first run 'rc.feature.filter.blanks' to remove non-sample derived signal.

Value

ramclustR object with total extracted ion normalized data.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

```
rc.feature.normalize.batch.qc  
rc.feature.normalize.batch.qc
```

Description

normalize data using batch.qc

Usage

```
rc.feature.normalize.batch.qc(  
  order = NULL,  
  batch = NULL,  
  qc = NULL,  
  ramclustObj = NULL,  
  qc.inj.range = 20  
)
```

Arguments

order	integer vector with length equal to number of injections in xset or csv file or dataframe
batch	integer vector with length equal to number of injections in xset or csv file or dataframe
qc	logical vector with length equal to number of injections in xset or csv file or dataframe
ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
qc.inj.range	integer: how many injections around each injection are to be scanned for presence of QC samples when using batch.qc normalization? A good rule of thumb is between 1 and 3 times the typical injection span between QC injections. i.e. if you inject QC ever 7 samples, set this to between 7 and 21. smaller values provide more local precision but make normalization sensitive to individual poor outliers (though these are first removed using the boxplot function outlier detection), while wider values provide less local precision in normalization but better stability to individual peak areas.

Value

ramclustR object with normalized data.

```
rc.feature.normalize.qc
      rc.feature.normalize.qc
```

Description

extractor for xcms objects in preparation for clustering

Usage

```
rc.feature.normalize.qc(
  ramclustObj = NULL,
  order = NULL,
  batch = NULL,
  qc.tag = NULL,
  output.plot = FALSE,
  p.cut = 0.05,
  rsq.cut = 0.1,
  p.adjust = "none"
)
```

Arguments

ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
order	integer vector with length equal to number of injections in xset or csv file
batch	integer vector with length equal to number of injections in xset or csv file
qc.tag	character vector of length one or two. If length is two, enter search string and factor name in \$phenoData slot (i.e. c("QC", "sample.type"). If length one (i.e. "QC"), will search for this string in the 'sample.names' slot by default.
output.plot	logical: if TRUE (default), plots are output to PDF.
p.cut	numeric when run order correction is applied, only features showing a run order vs signal with a linear p-value (after FDR correction) < p.cut will be adjusted. also requires r-squared < rsq.cut.
rsq.cut	numeric when run order correction is applied, only features showing a run order vs signal with a linear r-squared > rsq.cut will be adjusted. also requires p values < p.cut.
p.adjust	which p-value adjustment should be used? default = "none", see ?p.adjust

Details

This function offers normalization by run order, batch number, and QC sample signal intensity.

Each input vector should be the same length, and equal to the number of samples in the \$MSdata set.

Input vector order is assumed to be the same as the sample order in the \$MSdata set.

Value

ramclustR object with normalized data.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem.* 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

`rc.feature.normalize.quantile`
rc.feature.normalize.quantile

Description

normalize data using quantile

Usage

`rc.feature.normalize.quantile(ramclustObj = NULL)`

Arguments

`ramclustObj` ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)

Value

ramclustR object with normalized data.

`rc.feature.normalize.tic`
rc.feature.normalize.tic

Description

extractor for xcms objects in preparation for clustering

Usage

`rc.feature.normalize.tic(ramclustObj = NULL)`

Arguments

`ramclustObj` `ramclustObj` containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)

Details

This function offers normalization by total extracted ion signal. it is recommended to first run `'rc.feature.filter.blanks'` to remove non-sample derived signal.

Value

`ramclustR` object with total extracted ion normalized data.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem.* 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

`rc.feature.replace.na` *rc.feature.replace.na*

Description

replaces any NA (and optionally zero) values with small signal (20

Usage

```
rc.feature.replace.na(
  ramclustObj = NULL,
  replace.int = 0.1,
  replace.noise = 0.1,
  replace.zero = TRUE,
  which.data = c("MSdata", "MSMSdata")
)
```

Arguments

`ramclustObj` `ramclustObj` containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)

`replace.int` default = 0.1. proportion of minimum feature value to replace NA (or zero) values with

`replace.noise` default = 0.1. proportion of `replace.int` value by which noise is added via 'jitter'

`replace.zero` logical if TRUE, any zero values are replaced with noise as if they were NA values

`which.data` name of dataset

Details

noise is added by finding for each feature the minimum detected value, multiplying that value by `replace.int`, then adding (`replace.int*replace.noise`) noise. `abs()` is used to ensure no negative values result.

Value

ramclustR object with NA and zero values removed.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

rc.get.csv.data	<i>rc.get.csv.data</i>
-----------------	------------------------

Description

extractor for csv objects in preparation for normalization and clustering

Usage

```
rc.get.csv.data(  
  csv = NULL,  
  phenoData = NULL,  
  idmsms = NULL,  
  ExpDes = NULL,  
  sampNameCol = 1,  
  st = NULL,  
  timepos = 2,  
  featdelim = "_",  
  ensure.no.na = TRUE  
)
```

Arguments

csv	filepath: csv input. Features as columns, rows as samples. Column header <code>mz_rt</code>
phenoData	character: character string in 'taglocation' to designate files as either MS / DIA(MSe, MSall, AIF, etc) e.g. "01.mzML"
idmsms	filepath: optional idMSMS / MSe csv data. same dim and names as ms required

ExpDes	either an R object created by R ExpDes object: data used for record keeping and labelling msp spectral output
sampNameCol	integer: which column from the csv file contains sample names?
st	numeric: sigma t - time similarity decay value
timepos	integer: which position in delimited column header represents the retention time
featdelim	character: how feature mz and rt are delimited in csv import column header e.g. "="
ensure.no.na	logical: if TRUE, any 'NA' values in msint and/or msmsint are replaced with numerical values based on 10 percent of feature min plus noise. Used to ensure that spectra are not written with NA values.

Details

This function creates a ramclustObj which will be used as input for clustering.

Value

an empty ramclustR object. this object is formatted as an hclust object with additional slots for holding feature and compound data. details on these found below.

\$frt: feature retention time, in whatever units were fed in

\$fmz: feature retention time, reported in number of decimal points selected in ramclustR function

\$ExpDes: the experimental design object used when running ramclustR. List of two dataframes.

\$MSdata: the MSdataset provided by either xcms or csv input

\$MSMSdata: the (optional) DIA(MSe, MSall, AIF etc) dataset

\$xcmsOrd: original xcms order of features, for back-referencing when necessary

\$msint: weighted.mean intensity of feature in ms level data

\$msmsint: weighted.mean intensity of feature in msms level data

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. *Anal Chem*. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

Examples

```
## Choose csv input file. Features as columns, rows as samples
## Choose csv input file phenoData
filename <- system.file("extdata", "peaks.csv", package = "RAMClustR", mustWork = TRUE)
phenoData <- system.file("extdata", "phenoData.csv", package = "RAMClustR", mustWork = TRUE)

ramclustobj <- rc.get.csv.data(csv = filename, phenoData = phenoData, st = 5)
```

```
rc.get.df.data      rc.get.df.data
```

Description

extractor for dataframe input in preparation for normalization and clustering

Usage

```
rc.get.df.data(
  ms1_featureDefinitions = NULL,
  ms1_featureValues = NULL,
  ms2_featureDefinitions = NULL,
  ms2_featureValues = NULL,
  phenoData = NULL,
  ExpDes = NULL,
  featureNamesColumnIndex = 1,
  st = NULL,
  ensure.no.na = TRUE
)
```

Arguments

ms1_featureDefinitions	dataframe with metadata with columns: mz, rt, feature names containing MS data
ms1_featureValues	dataframe with rownames = sample names, colnames = feature names containing MS data
ms2_featureDefinitions	dataframe with metadata with columns: mz, rt, feature names containing MSMS data
ms2_featureValues	dataframe with rownames = sample names, colnames = feature names containing MSMS data
phenoData	dataframe containing phenoData
ExpDes	either an R object created by R ExpDes object: data used for record keeping and labelling msp spectral output

featureNamesColumnIndex	integer: which column in 'ms1_featureDefinitions' contains feature names?
st	numeric: sigma t - time similarity decay value
ensure.no.na	logical: if TRUE, any 'NA' values in msint and/or msmsint are replaced with numerical values based on 10 percent of feature min plus noise. Used to ensure that spectra are not written with NA values.

Details

This function creates a ramclustObj which will be used as input for clustering.

Value

an empty ramclustR object. this object is formatted as an hclust object with additional slots for holding feature and compound data. details on these found below.

\$frt: feature retention time, in whatever units were fed in

\$fmz: feature retention time, reported in number of decimal points selected in ramclustR function

\$ExpDes: the experimental design object used when running ramclustR. List of two dataframes.

\$MSdata: the MSdataset provided by either xcms or csv input

\$MSMSdata: the (optional) DIA(MSe, MSall, AIF etc) dataset

\$xcmsOrd: original xcms order of features, for back-referencing when necessary

\$msint: weighted.mean intensity of feature in ms level data

\$msmsint: weighted.mean intensity of feature in msms level data

Author(s)

Zargham Ahmad, Helge Hecht, Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem.* 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. *Anal Chem.* 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

Examples

```
## Choose dataframe with metadata with columns: mz, rt, feature names containing MS data
## Choose dataframe with rownames = sample names, colnames = feature names containing MS data
## Choose dataframe containing phenoData
df1 <- readRDS(system.file("extdata", "featDefinition.rds", package = "RAMClustR", mustWork = TRUE))
df2 <- readRDS(system.file("extdata", "featValues.rds", package = "RAMClustR", mustWork = TRUE))
df3 <- readRDS(system.file("extdata", "phenoData_df.rds", package = "RAMClustR", mustWork = TRUE))
```

```
ramclustr <- rc.get.df.data(ms1_featureDefinitions=df1, ms1_featureValues=df2, phenoData=df3, st=5)
```

```
rc.get.xcms.data      rc.get.xcms.data
```

Description

extractor for xcms objects in preparation for normalization and clustering

Usage

```
rc.get.xcms.data(
  xcmsObj = NULL,
  taglocation = "filepaths",
  MStag = NULL,
  MSMStag = NULL,
  ExpDes = NULL,
  mzdec = 3,
  ensure.no.na = TRUE
)
```

Arguments

xcmsObj	xcmsObject: containing grouped feature data for clustering by ramclustr
taglocation	character: "filepaths" by default, "phenoData[,1]" is another option. refers to xcms slot
MStag	character: character string in 'taglocation' to designate files as either MS / DIA(MSe, MSall, AIF, etc) e.g. "01.mzML"
MSMStag	character: character string in 'taglocation' to designate files as either MS / DIA(MSe, MSall, AIF, etc) e.g. "02.mzML"
ExpDes	either an R object created by R ExpDes object: data used for record keeping and labelling msp spectral output
mzdec	integer: number of decimal places for storing m/z values
ensure.no.na	logical: if TRUE, any 'NA' values in msint and/or msmsint are replaced with numerical values based on 10 percent of feature min plus noise. Used to ensure that spectra are not written with NA values.

Details

This function creates a ramclustrObj which will be used as input for clustering.

Value

an empty ramclustR object. this object is formatted as an hclust object with additional slots for holding feature and compound data. details on these found below.

\$frt: feature retention time, in whatever units were fed in (xcms uses seconds, by default)

\$fmz: feature retention time, reported in number of decimal points selected in ramclustR function

\$ExpDes: the experimental design object used when running ramclustR. List of two dataframes.

\$MSdata: the MSdataset provided by either xcms or csv input

\$MSMSdata: the (optional) DIA(MSe, MSall, AIF etc) dataset provided by either xcms or csv input

\$xcmsOrd: original xcms order of features, for back-referencing when necessary

\$msint: weighted.mean intensity of feature in ms level data

\$msmsint: weighted.mean intensity of feature in msms level data

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. *Anal Chem*. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

rc.merge.split.clusters

rc.merge.split.clusters

Description

Cluster refinement - scanning instruments (quadrupole, as in GC-MS) can display cluster splitting, possibly due to slight differences in measured peak retention time as a function of mass due to scan dynamics. this function enables a second pass clustering designed to merge two clusters if the second cluster is within a small retention time window and shows a sufficiently strong correlation.

Usage

```
rc.merge.split.clusters(  
  ramclustObj = NULL,  
  merge.threshold = 0.7,  
  cor.method = "spearman",  
  rt.sd.factor = 3  
)
```

Arguments

<code>ramclustObj</code>	ramclustR object to annotate.
<code>merge.threshold</code>	numeric. value between -1 and 1 indicating the correlational r threshold above which two clusters will be merged
<code>cor.method</code>	character. default = 'spearman'. correlational method to use for calculating r. see documentation on R base <code>cor()</code> function for available options
<code>rt.sd.factor</code>	numeric. default = 3. clusters within <code>rt.sd.factor * ramclustObj\$rtstd</code> (cluster retention time standard deviation) are considered for merging.

Details

exports files to a directory called 'spectra'. If `one.file = FALSE`, a new directory 'spectra/msp' is created to hold the individual msp files. if `do.findman` has been run, spectra are written as `ms2` spectra, else as `ms1`.

Value

new ramclustR object, with (generally) fewer clusters than the input ramclustR object.

Author(s)

Corey Broeckling

rc.qc

rc.qc

Description

summarize quality control for clustering and for quality control sample variation based on compound (`$SpecAbund`) and feature (`$MSdata` and `$MSMSdata`, if present)

Usage

```
rc.qc(  
  ramclustObj = NULL,  
  qc.tag = "QC",  
  remove.qc = FALSE,  
  npc = 4,  
  scale = "pareto",  
  outfile.basename = "ramclustQC",  
  view.hist = TRUE  
)
```

Arguments

ramclustObj	ramclustR object to analyze
qc.tag	qc.tag character vector of length one or two. If length is two, enter search string and factor name in \$phenoData slot (i.e. c("QC", "sample.type"). If length one (i.e. "QC"), will search for this string in the 'sample.names' slot by default.
remove.qc	logical - if TRUE (default) QC injections will be removed from the returned ramclustObj (applies to \$MSdata, \$MSMSdata, \$SpecAbund, \$phenoData, as appropriate). If FALSE, QC samples remain.
npc	number of Principle components to calculate and plot
scale	"pareto" by default: PCA scaling method used
outfile.basename	base name of output files. Extensions added internally. default = "ramclustQC"
view.hist	logical. should histograms be plotted?

Details

plots a ramclustR summary plot. first page represents the correlation of each cluster to all other clusters, sorted by retention time. large blocks of yellow along the diagonal indicate either poor clustering or a group of coregulated metabolites with similar retention time. It is an imperfect diagnostic, particularly with lipids on reverse phase LC or sugars on HILIC LC systems. Page 2: histogram of r values from page 1 - only r values one position from the diagonal are used. Pages 3:5 - PCA results, with QC samples colored red. relative standard deviation calculated as $sd(QC\ PC\ scores) / sd(all\ PC\ scores)$. Page 6: histogram of CV values for each compound in the dataset, QC samples only.

Value

new RC object. Saves output summary plots to pdf and .csv summary tables to new 'QC' directory. If remove.qc = TRUE, moves QC samples to new \$QC slot from original position.

Author(s)

Corey Broeckling

References

- Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.
- Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. *Anal Chem*. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

rc.ramclustr	<i>rc.ramclustr</i>
--------------	---------------------

Description

Main clustering function for grouping features based on their analytical behavior.

Usage

```
rc.ramclustr(
  ramclustObj = NULL,
  st = NULL,
  sr = NULL,
  maxt = NULL,
  deepSplit = FALSE,
  blocksize = 2000,
  mult = 5,
  hmax = NULL,
  collapse = TRUE,
  minModuleSize = 2,
  linkage = "average",
  cor.method = "pearson",
  rt.only.low.n = TRUE
)
```

Arguments

ramclustObj	ramclustR object: containing ungrouped features. constructed by rc.get.xcms.data, for example
st	numeric: sigma t - time similarity decay value
sr	numeric: sigma r - correlational similarity decay value
maxt	numeric: maximum time difference to calculate retention similarity for - all values beyond this are assigned similarity of zero
deepSplit	logical: controls how aggressively the HCA tree is cut - see ?cutreeDynamicTree
blocksize	integer: number of features (scans?) processed in one block =1000,
mult	numeric: internal value, can be used to influence processing speed/ram usage
hmax	numeric: precut the tree at this height, default 0.3 - see ?cutreeDynamicTree
collapse	logical: if true (default), feature quantitative values are collapsed into spectra quantitative values.
minModuleSize	integer: how many features must be part of a cluster to be returned? default = 2
linkage	character: heirarchical clustering linkage method - see ?hclust
cor.method	character: which correlational method used to calculate 'r' - see ?cor
rt.only.low.n	logical: default = TRUE At low injection numbers, correlational relationships of peak intensities may be unreliable. by default ramclustR will simply ignore the correlational r value and cluster on retention time alone. if you wish to use correlation with at n < 5, set this value to FALSE.

Details

Main clustering function output - see citation for algorithm description or vignette('RAMClustR') for a walk through. batch.qc. normalization requires input of three vectors (1) batch (2) order (3) qc. This is a feature centric normalization approach which adjusts signal intensities first by comparing batch median intensity of each feature (one feature at a time) QC signal intensity to full dataset median to correct for systematic batch effects and then secondly to apply a local QC median vs global median sample correction to correct for run order effects.

Value

\$featclus: integer vector of cluster membership for each feature

\$clrt: cluster retention time

\$clrtsd: retention time standard deviation of all the features that comprise that cluster

\$nfeat: number of features in the cluster

\$nsing: number of 'singletons' - that is the number of features which clustered with no other feature

\$cmpd: compound name. C#### are assigned in order of output by dynamicTreeCut. Compound with the most features is classified as C0001...

\$ann: annotation. By default, annotation names are identical to 'cmpd' names. This slot is a placeholder for when annotations are provided

\$SpecAbund: the cluster intensities after collapsing features to clusters

\$SpecAbundAve: the cluster intensities after averaging all samples with identical sample names

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. *Anal Chem*. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

rc.remove.qc

rc.remove.qc

Description

summarize quality control for clustering and for quality control sample variation based on compound (\$SpecAbund) and feature (\$MSdata and \$MSMSdata, if present)

Usage

```
rc.remove.qc(ramclustObj = NULL, qc.tag = "QC")
```

Arguments

ramclustObj ramclustR object to analyze

qc.tag qc.tag character vector of length one or two. If length is two, enter search string and factor name in \$phenoData slot (i.e. c("QC", "sample.type"). If length one (i.e. "QC"), will search for this string in the 'sample.names' slot by default.

Details

simply moves QC samples out of the way for downstream processing. moved to a \$qc slot.

Value

new RC object. moves QC samples to new \$qc slot from original position.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. *Anal Chem*. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

rc.restore.qc.samples *rc.restore.qc.samples*

Description

summarize quality control for clustering and for quality control sample variation based on compound (\$SpecAbund) and feature (\$MSdata and \$MSMSdata, if present)

Usage

```
rc.restore.qc.samples(ramclustObj = NULL)
```

Arguments

ramclustObj ramclustR object to analyze

Details

moves all of \$phenoData, \$MSdata, \$MSMSdata, \$SpecAbund back to original positions from \$qc slot

Value

RC object

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. *Anal Chem*. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

RCQC

RCQC

Description

filter RC object and summarize quality control sample variation

Usage

```
RCQC(
  ramclustObj = NULL,
  qctag = "QC",
  npc = 4,
  scale = "pareto",
  which.data = "SpecAbund",
  outfile = "ramclustQC.pdf"
)
```

Arguments

ramclustObj	ramclustR object to analyze
qctag	"QC" by default - rowname tag to identify QC samples
npc	number of Principle components to calculate and plot
scale	"pareto" by default: PCA scaling method used
which.data	which dataset to use. "SpecAbund" by default
outfile	name of output pdf file.

Details

plots a ramclustR summary plot. first page represents the correlation of each cluster to all other clusters, sorted by retention time. large blocks of yellow along the diagonal indicate either poor clustering or a group of coregulated metabolites with similar retention time. It is an imperfect diagnostic, particularly with lipids on reverse phase LC or sugars on HILIC LC systems. Page 2: histogram of r values from page 1 - only r values one position from the diagonal are used. Pages 3:5 - PCA results, with QC samples colored red. relative standard deviation calculated as $\text{sd}(\text{QC PC scores}) / \text{sd}(\text{all PC scores})$. Page 6: histogram of CV values for each compound in the dataset, QC samples only.

Value

new RC object, with QC samples moved to new slot. prints output summary plots to pdf.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. *Anal Chem*. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

remove_blanks

remove_blanks

Description

remove blanks

Usage

```
remove_blanks(ramclustObj, blank)
```

Arguments

ramclustObj	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
blank	blank samples found by define_samples

Value

ramclustObj object with blanks removed

replace_na	<i>replace_na</i>
------------	-------------------

Description

add rc.feature.replace.na params in ramclustObj

Usage

```
replace_na(data, replace.int, replace.zero, replace.noise)
```

Arguments

data	selected data frame to use
replace.int	default = 0.1. proportion of minimum feature value to replace NA (or zero) values with
replace.zero	logical if TRUE, any zero values are replaced with noise as if they were NA values
replace.noise	default = 0.1. proportion of replace.int value by which noise is added via 'jitter'

Value

selected ramclustR data frame with NA and zero values removed.
number of features replaced

write.gcei.mat	<i>write.gcei.mat</i>
----------------	-----------------------

Description

Export GC-MS EI spectra for spectral searching in MSFinder

Usage

```
write.gcei.mat(ramclustObj = NULL)
```

Arguments

ramclustObj	ramclustR object to annotate.
-------------	-------------------------------

Details

exports files to a directory called 'spectra'. a new directory 'spectra/mat' is created to hold the individual mat files.

Value

nothing, just exports files to the working directory

Author(s)

Corey Broeckling

write.methods	<i>write.methods</i>
---------------	----------------------

Description

write RAMClustR processing methods and citations to text file

Usage

```
write.methods(ramclustObj = NULL, filename = NULL)
```

Arguments

ramclustObj	R object - the ramclustR object which was used to write the .mat or .msp files
filename	define filename/path to write. uses 'ramclustr_methods.txt' and the working directory by default.

Details

this function exports a file called ramclustr_methods.txt which contains the processing history, parameters used, and relevant citations.

Value

an annotated ramclustR object
nothing - new file written to working director

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem*. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

`write.msp`*write.msp*

Description

Cluster annotation function: inference of 'M' - molecular weight of the compound giving rise to each spectrum - using the InterpretMSSpectrum::findMain function

Usage

```
write.msp(ramclustObj = NULL, one.file = FALSE)
```

Arguments

<code>ramclustObj</code>	ramclustR object to annotate.
<code>one.file</code>	logical, should all msp spectra be written to one file? If false, each spectrum is an individual file.

Details

exports files to a directory called 'spectra'. If `one.file = FALSE`, a new directory 'spectra/msp' is created to hold the individual msp files. if `do.findman` has been run, spectra are written as ms2 spectra, else as ms1.

Value

nothing, just exports files to the working directory

Author(s)

Corey Broeckling

`write_csv`*write_csv*

Description

write csv template called "ExpDes.csv" to your working directory. you will fill this in manually, ensuring that when you save you retain csv format. ramclustR will then read this file in and format appropriately.

Usage

```
write_csv(data)
```

write_csv

55

Arguments

data csv template to write

Value

read ExpDes.csv file

Index

* MSFinder

- defineExperiment, 8
- findfeature, 14
- findmass, 15
- ramclustR, 23
- rc.calibrate.ri, 26
- rc.cmpd.filter.blanks, 27
- rc.cmpd.filter.cv, 29
- rc.cmpd.replace.na, 30
- rc.expand.sample.names, 31
- rc.feature.filter.blanks, 32
- rc.feature.filter.cv, 34
- rc.feature.normalize.qc, 36
- rc.feature.normalize.tic, 37
- rc.feature.replace.na, 38
- rc.get.csv.data, 39
- rc.get.df.data, 41
- rc.get.xcms.data, 43
- rc.qc, 45
- rc.ramclustr, 47
- rc.remove.qc, 48
- rc.restore.qc.samples, 49
- RCQC, 50

* RAMClustR

- defineExperiment, 8
- doFindmain, 9
- exportDataset, 12
- findfeature, 14
- findmass, 15
- getData, 17
- mergeRObjects, 19
- ramclustR, 23
- rc.calibrate.ri, 26
- rc.cmpd.filter.blanks, 27
- rc.cmpd.filter.cv, 29
- rc.cmpd.replace.na, 30
- rc.expand.sample.names, 31
- rc.export.msp.rc, 32
- rc.feature.filter.blanks, 32

- rc.feature.filter.cv, 34
- rc.feature.normalize.qc, 36
- rc.feature.normalize.tic, 37
- rc.feature.replace.na, 38
- rc.get.csv.data, 39
- rc.get.df.data, 41
- rc.get.xcms.data, 43
- rc.merge.split.clusters, 44
- rc.qc, 45
- rc.ramclustr, 47
- rc.remove.qc, 48
- rc.restore.qc.samples, 49
- RCQC, 50
- write.gcei.mat, 52
- write.methods, 53
- write.msp, 54

* clustering

- defineExperiment, 8
- doFindmain, 9
- findfeature, 14
- findmass, 15
- mergeRObjects, 19
- ramclustR, 23
- rc.calibrate.ri, 26
- rc.cmpd.filter.blanks, 27
- rc.cmpd.filter.cv, 29
- rc.cmpd.replace.na, 30
- rc.expand.sample.names, 31
- rc.export.msp.rc, 32
- rc.feature.filter.blanks, 32
- rc.feature.filter.cv, 34
- rc.feature.normalize.qc, 36
- rc.feature.normalize.tic, 37
- rc.feature.replace.na, 38
- rc.get.csv.data, 39
- rc.get.df.data, 41
- rc.get.xcms.data, 43
- rc.merge.split.clusters, 44
- rc.qc, 45

- rc.ramclustr, 47
- rc.remove.qc, 48
- rc.restore.qc.samples, 49
- RCQC, 50
- write.gcei.mat, 52
- write.methods, 53
- write.msp, 54
- * **feature**
 - defineExperiment, 8
 - doFindmain, 9
 - findfeature, 14
 - findmass, 15
 - mergeRObjects, 19
 - ramclustR, 23
 - rc.calibrate.ri, 26
 - rc.cmpd.filter.blanks, 27
 - rc.cmpd.filter.cv, 29
 - rc.cmpd.replace.na, 30
 - rc.expand.sample.names, 31
 - rc.feature.filter.blanks, 32
 - rc.feature.filter.cv, 34
 - rc.feature.normalize.qc, 36
 - rc.feature.normalize.tic, 37
 - rc.feature.replace.na, 38
 - rc.get.csv.data, 39
 - rc.get.df.data, 41
 - rc.get.xcms.data, 43
 - rc.qc, 45
 - rc.ramclustr, 47
 - rc.remove.qc, 48
 - rc.restore.qc.samples, 49
 - RCQC, 50
 - write.gcei.mat, 52
 - write.methods, 53
 - write.msp, 54
- * **findMain**
 - doFindmain, 9
- * **interpretMSSpectrum**
 - doFindmain, 9
 - rc.export.msp.rc, 32
 - write.gcei.mat, 52
 - write.msp, 54
- * **mass spectrometry**
 - defineExperiment, 8
 - doFindmain, 9
 - findfeature, 14
 - findmass, 15
 - mergeRObjects, 19
 - ramclustR, 23
 - rc.calibrate.ri, 26
 - rc.cmpd.filter.blanks, 27
 - rc.cmpd.filter.cv, 29
 - rc.cmpd.replace.na, 30
 - rc.expand.sample.names, 31
 - rc.export.msp.rc, 32
 - rc.feature.filter.blanks, 32
 - rc.feature.filter.cv, 34
 - rc.feature.normalize.qc, 36
 - rc.feature.normalize.tic, 37
 - rc.feature.replace.na, 38
 - rc.get.csv.data, 39
 - rc.get.df.data, 41
 - rc.get.xcms.data, 43
 - rc.merge.split.clusters, 44
 - rc.qc, 45
 - rc.ramclustr, 47
 - rc.remove.qc, 48
 - rc.restore.qc.samples, 49

- RCQC, 50
- write.gcei.mat, 52
- write.methods, 53
- write.msp, 54
- * **ramclustR**
 - doFindmain, 9
 - exportDataset, 12
 - findfeature, 14
 - findmass, 15
 - mergeRObjects, 19
 - ramclustR, 23
 - rc.calibrate.ri, 26
 - rc.cmpd.filter.blanks, 27
 - rc.cmpd.filter.cv, 29
 - rc.cmpd.replace.na, 30
 - rc.expand.sample.names, 31
 - rc.export.msp.rc, 32
 - rc.feature.filter.blanks, 32
 - rc.feature.filter.cv, 34
 - rc.feature.normalize.qc, 36
 - rc.feature.normalize.tic, 37
 - rc.feature.replace.na, 38
 - rc.get.csv.data, 39
 - rc.get.df.data, 41
 - rc.get.xcms.data, 43
 - rc.merge.split.clusters, 44
 - rc.qc, 45
 - rc.ramclustr, 47
 - rc.remove.qc, 48
 - rc.restore.qc.samples, 49
 - RCQC, 50
 - write.gcei.mat, 52
 - write.methods, 53
 - write.msp, 54
- * **ramclustR**
 - defineExperiment, 8
- * **xcms**
 - defineExperiment, 8
 - doFindmain, 9
 - findfeature, 14
 - findmass, 15
 - ramclustR, 23
 - rc.calibrate.ri, 26
 - rc.cmpd.filter.blanks, 27
 - rc.cmpd.filter.cv, 29
 - rc.cmpd.replace.na, 30
 - rc.expand.sample.names, 31
 - rc.export.msp.rc, 32
 - rc.feature.filter.blanks, 32
 - rc.feature.filter.cv, 34
 - rc.feature.normalize.qc, 36
 - rc.feature.normalize.tic, 37
 - rc.feature.replace.na, 38
 - rc.get.csv.data, 39
 - rc.get.df.data, 41
 - rc.get.xcms.data, 43
 - rc.merge.split.clusters, 44
 - rc.qc, 45
 - rc.ramclustr, 47
 - rc.remove.qc, 48
 - rc.restore.qc.samples, 49
 - RCQC, 50
 - write.gcei.mat, 52
 - write.methods, 53
 - write.msp, 54
- add_params, 3
- check_arguments_filter.blanks, 4
- check_arguments_filter.cv, 4
- check_arguments_replace.na, 5
- checks, 3
- compute_do.sets, 5
- compute_SpecAbundAve, 6
- compute_wt_mean, 6
- create_ramclustObj, 7
- define_samples, 9
- defineExperiment, 8
- doFindmain, 9
- exportDataset, 12
- filter_blanks, 13
- filter_good_features, 13
- filter_signal, 14
- find_good_features, 16
- findfeature, 14
- findmass, 15
- get_ExpDes, 18
- get_instrument_platform, 18
- getData, 17
- mean_signal_intensities, 19
- mergeRObjects, 19
- normalized_data_batch_qc, 21

normalized_data_tic, 22

order_datasets, 22

ramclustR, 23

rc.calibrate.ri, 26

rc.cmpd.filter.blanks, 27

rc.cmpd.filter.cv, 29

rc.cmpd.replace.na, 30

rc.expand.sample.names, 31

rc.export.msp.rc, 32

rc.feature.filter.blanks, 32

rc.feature.filter.cv, 34

rc.feature.normalize.batch.qc, 35

rc.feature.normalize.qc, 36

rc.feature.normalize.quantile, 37

rc.feature.normalize.tic, 37

rc.feature.replace.na, 38

rc.get.csv.data, 39

rc.get.df.data, 41

rc.get.xcms.data, 43

rc.merge.split.clusters, 44

rc.qc, 45

rc.ramclustr, 47

rc.remove.qc, 48

rc.restore.qc.samples, 49

RCQC, 50

remove_blanks, 51

replace_na, 52

write.gcei.mat, 52

write.methods, 53

write.msp, 54

write_csv, 54