Package 'RAMClustR'

May 12, 2021

Type Package

Title Mass Spectrometry Metabolomics Feature Clustering and Interpretation

Version 1.2.1

Date 2021-05-11

Author Corey D. Broeckling, Fayyaz Afsar, Steffan Neumann, Asa Ben-Hur, Jessica Prenni.

Maintainer ``Broeckling,Corey" <Corey.Broeckling@ColoState.EDU>

Imports dynamicTreeCut, fastcluster, ff, InterpretMSSpectrum, BiocManager, httr, jsonlite, preprocessCore, e1071, gplots, pcaMethods, stringr, xml2, utils, webchem, stringi, RCurl, MSnbase, ggplot2

License GPL (>= 2)

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 7.1.1

Suggests knitr, rmarkdown, xcms

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2021-05-12 05:52:15 UTC

R topics documented:

annotate *evalute ramSearch, MSFinder mssearch, MSFinder Structure, MSFinder Formula, and findmain output to annotate spectra of ramclustR object*

Description

After running RAMSearch (msp) and MSFinder on .mat or .msp files, import the spectral search results

Usage

```
annotate(
  ramclustObj = NULL,
  standardize.names = FALSE,
  min.msms.score = 6.5,
  database.priority = NULL,
  database.priority.factor = 0.1,
  find.inchikey = TRUE,
  taxonomy.inchi = NULL,
  taxonomy.inchi.factor = 0.1,
  use.ri = TRUE,
  sri = 300,
  ri.na.factor = 0.6,
  reset = TRUE
)
```

Arguments

`ramclustObj`	R object - the ramclustR object which was used to write the .mat or .msp files
`standardize.names`	logical: if TRUE, use inchikey for standardized chemical name lookup (http://cts.fiehnlab.ucdavis.edu/)
`min.msms.score`	numerical: what is the minimum MSFinder similarity score acceptable. default = 6.5
`database.priority`	character. Formula assignment prioritization based on presence in one or more (structure) databases. Can be set to a single or multiple database names. must match database names as they are listed in MSFinder precisily. Can also be set to 'all' (note that MSFinder reports all databases matched, not just databases in MSFinder parameters). If any database is set, the best formula match to any of those databases is selected, rather than the best formula match overall. If NULL, this will be set to include all selected databases (from ramclustObj$msfinder.dbs, retreieved from search output during import.msfinder.formulas(), when available) or 'all'.
`database.priority.factor`	numeric, between 0 and 1. 0.1 by default. The proportion by which scores for structures not in priority database are assessed

find.inchikey	logical. default = TRUE. use chemical translation service to try to look up inchikey for chemical name.
taxonomy.inchi	vector or data frame. Only when rescore.structure = TRUE. user can supply a vector of inchikeys. If used, structures which match first block of inchikey retain full score, while all other structures are penalized.
taxonomy.inchi.factor	numeric, between 0 and 1. 0.1 by default. The proportion by which scores for structures not in taxonomy.inchi vector are assessed
use.ri	logical. default = TRUE. If retention index is available in ramclustObj (set by 'rc.calibrate.ri') and in library spectra from MSFinder, use RI similiarity to rescore.
sri	numeric. sigma value for retention index. controls decay rate of retention index curve. decay rate between 0 and 1 exported, and multiplied by spectrum score, totalscore.
ri.na.factor	numeric. between 0 and 1. 0.5 by default. how should spectrum scores be treated when no retention index is available? NA values are replaced by retention index similarities of ri.na.factor when use.ri = TRUE.
reset	logical. If TRUE, removes any previously assigned annotations.

Details

this function imports the output from the MSFinder program to annotate the ramclustR object

Value

an updated ramclustR object, with the at $msfinder.formula, $msfinder.formula.score, $ann, and $ann.conf slots updated to annotated based on output from 1. ramsearch output, 2. msfinder mssearch, 3. msfinder predicted structure, 4. msfinder predicted formula, and 5. interpretMSSpectrum inferred molecular weight, with listed order as priority.

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.

Tsugawa H, Kind T, Nakabayashi R, Yukihira D, Tanaka W, Cajka T, Saito K, Fiehn O, Arita M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem. 2016 Aug 16;88(16):7946-58. doi: 10.1021/acs.analchem.6b00770. Epub 2016 Aug 4. PubMed PMID: 27419259.

http://cts.fiehnlab.ucdavis.edu/static/download/CTS2-MS2015.pdf

annotation.summary *annotation.summary()*

Description

Write a .csv file containing a summary of the annotations in the ramclustR object.

Usage

```
annotation.summary(ramclustObj = NULL, outfile = NULL)
```

Arguments

ramclustObj	R object - the ramclustR object which was used to write the .mat or .msp files
outfile	file path/name of output csv summary file. if NULL (default) will be exported to spectra/annotaionSummary.csv

Details

this function exports a csv file summarizing annotation evidence for each compound

Value

nothing

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.

assign.z *assign.z*

Description

infer charge state of features in ramclustR object.

Usage

```
assign.z(
  ramclustObj = NULL,
  chargestate = c(1:5),
  mzError = 0.02,
  nEvents = 2,
  minPercentSignal = 10,
  assume1 = TRUE
)
```

Arguments

ramclustObj	ramclustR object to annotate
chargestate	integer vector. vector of integers of charge states to look for. default = c(1:5)
mzError	numeric. the error allowed in charge state m/z filtering. absolute mass units
nEvents	integer. the number of isotopes necessary to assign a charnge state > 1. default = 2.
minPercentSignal	numeric. the ratio of isotope signal (all isotopes) divided by total spectrum signal * 100 much be greated than minPercentSignal to evaluate charge state. Value should be between 0 and 100.
assume1	logical. when TRUE, m/z values for which no isotopes are found are assumed to be at z = 1.

Details

Annotation of ramclustR spectra. looks at isotope spacing for clustered features to infer charge state for each feature and a max charge state for each compound

Value

returns a ramclustR object. new slots holding:

zmax. vector with length equal to number of compounds. max charge state detected for that compound

fm. vector of inferred 'm', m/z value * z value

fz. vector of inferred 'z' values based on analysis of isotopes in spectrum.

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.

`change.annotation` *evalute ramSearch, MSFinder mssearch, MSFinder Structure, MSFinder Formula, and findmain output to annotate spectra of ramclustR object*

Description

After running RAMSearch (msp) and MSFinder on .mat or .msp files, import the spectral search results

Usage

```
change.annotation(
  ramclustObj = NULL,
  msfinder.dir = "C:/MSFinder/MSFINDER ver 3.22",
  standardize.names = FALSE,
  min.msms.score = 3.5,
  database.priority = "all",
  any.database.priority = TRUE,
  reset = TRUE
)
```

Arguments

`ramclustObj`	R object - the ramclustR object which was used to write the .mat or .msp files
`msfinder.dir`	full path to MSFinder directory - used for naming refinement
`standardize.names`	logical: if TRUE, use inchikey for standardized chemical name lookup (http://cts.fiehnlab.ucdavis.edu/)
`min.msms.score`	numerical: what is the minimum MSFinder similarity score acceptable. default = 3.5
`database.priority`	character. Formula assignment prioritization based on presence in one or more databases. Can be set to a single or multiple database names. must match database names as they are listed in MSFinder precisily. Can also be set to 'all' (note that MSFinder reports all databases matched, not just selected databases). If any database is set, the best formula match to that (those) database(s) is selected, rather than the best formula match overall.
`any.database.priority`	logical. First priority in formula assignment is based on any of the 'database.priority' values. Secondary priority from all other databases (determined in original MSFinder search) if TRUE. If false, formula assignment score from MSFinder used independent of structure search results.
`reset`	logical. If TRUE, removes any previously assigned annotations.

Details

this function imports the output from the MSFinder program to annotate the ramclustR object

Value

an updated ramclustR object, with the at $msfinder.formula, $msfinder.formula.score, $ann, and $ann.conf slots updated to annotated based on output from 1. ramsearch output, 2. msfinder mssearch, 3. msfinder predicted structure, 4. msfinder predicted formula, and 5. interpretMSSpectrum inferred molecular weight, with listed order as priority.

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.

Tsugawa H, Kind T, Nakabayashi R, Yukihira D, Tanaka W, Cajka T, Saito K, Fiehn O, Arita M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem. 2016 Aug 16;88(16):7946-58. doi: 10.1021/acs.analchem.6b00770. Epub 2016 Aug 4. PubMed PMID: 27419259.

http://cts.fiehnlab.ucdavis.edu/static/download/CTS2-MS2015.pdf

---

`cmpd.summary` *cmpd.summary*

Description

a bit of reporting for compounds, quick access summary and plot (if available)

Usage

```
cmpd.summary(ramclustObj = NULL, cmpd = 1)
```

Arguments

`ramclustObj`	ramclustR object to annotate
`cmpd`	integer. compound number to report. i.e. 459.

Details

Reports name, annotation, retention time, number of features in spectrum, median and mean signal intensity, and if interpretMSSpectrum (do.findmain) has been run, plots an annotated MS level spectrum.

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.

`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 leaset) 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.

`do.findmain` *do.findmain*

Description

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

Usage

```
do.findmain(
  ramclustObj = NULL,
  cmpd = 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.
`cmpd`	integer: vector defining compound numbers to annotated. if NULL (default), all compounds
`mode`	character: "positive" or "negative"
`mzabs.error`	numeric: absolute mass deviation allowd, default = 0.01

ppm.error	numeric: ppm mass error _added_ to mzabs.error, default = 10
ads	character: vector of allowed adducts, i.e. c("[M+H]+"). if NULL, default positive mode values of H+, Na+, K+, and NH4+, as monomer, dimer, and trimer, are assigned. Negative mode include "[M-H]-", "[M+Na-2H]-", "[M+K-2H]-", "[M+CH2O2-H]-" as monomer, dimer, and trimer.
nls	character: vector of allowed neutral losses, i.e. c("[M+H-H2O]+"). 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 polts be generated for evaluation? detfault = TRUE. PDF saved to working.directory/spectra
writeMat	logical: should indidual .mat files (for MSFinder) be generated in a 'mat' subdirectory in the 'spectra' folder? default = TRUE.
writeMS	logical: should indidual .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 exporing spectra for annotation.

Details

a partially annotated ramclustR object. base structure is that of a standard R heirarchical 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 intepretation 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 concflicts between ramclustR and findmain M assignment when scoring = auto.

$M.ann.ramclustr: annotated spectrum supporing M using ramclustR scoring

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

$M.ann.findmain: annotated spectrum supporing M using findmain scoring

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

Author(s)

Corey Broeckling

References

Jaeger C, ... Lisec 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.

---

export.msfinder.formulas

export MSFinder formula prediction results in tabular format.

Description

After running MSFinder, results have been imported to the ramclustR object. This function exports as a .csv file for ease of viewing.

Usage

```
export.msfinder.formulas(
  ramclustObj = NULL,
  export.all = FALSE,
  output.directory = NULL
)
```

Arguments

`ramclustObj`	R object - the ramclustR object which was used to write the .mat or .msp files
`export.all`	logical: default = FALSE. If TRUE, export all columns, if FALSE, only columns 1: "exactmass"
`output.directory`	valid path: default = NULL. If NULL, results are exported to spectra/mat directory.

Details

this function exports a .csv file containing all returned MSFinder molecular formula hypotheses. this file is saved (by default) to the working directory spectra/mat/ directory

Value

an updated ramclustR object, with the RC$ann and RC$ann.conf slots updated to annotated based on output from 1. ramsearch output, 2. msfinder mssearch, 3. msfinder predicted structure, 4. msfinder predicted formula, and 5. interpretMSSpectrum inferred molecular weight, with listed order as priority.

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.

Tsugawa H, Kind T, Nakabayashi R, Yukihira D, Tanaka W, Cajka T, Saito K, Fiehn O, Arita M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem. 2016 Aug 16;88(16):7946-58. doi: 10.1021/acs.analchem.6b00770. Epub 2016 Aug 4. PubMed PMID: 27419259.

`exportDataset` *exportDataset*

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",
  filter = 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)
filter	logical, TRUE by default. when $cmpd.use slot is present (from rc.cmpd.filter.cv function), only cmpds that passed cv filtering are used. If you wish to change that behavior, rerun the rc.cmpd.filter.cv function with a really high CV threshold.

Details

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

Value

nothing is returned. file exported as csf 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.

---

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, feaure 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 isottopic 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

---

fooddb2msfinder *foodb2msfinder*

Description

convenience function for converting FoodDB database export format to MSFinder custom database import format. Before running this, please have downloaded .csv files from FoodDB with the appropriate Display Field Headers (see details)

Usage

```
fooddb2msfinder(
  foodb.files = NULL,
  out.dir = NULL,
  out.name = "FoodDB_for_MSFinder.txt"
)
```

Arguments

foodb.files	default = NULL, if path is set, will read automatically. If NULL, direcory selection by user.
out.dir	default = NULL. Can set to exiseting directory with full path name. If NULL, direcory selection by user.
out.name	default = "FoodDB_for_MSFinder.txt".

Details

Input file(s) should be csv formatted, with required headers of 'Name', 'Smiles', 'Inchikey', 'Chemical formula', and 'Mono mass' - case sensitive. Ouput will be in tab delimited text format in directory of choice.

Value

Nothing is returned - output file written to directory set by 'out.dir' and name set by 'out.name'

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.

---

getData *getData*

Description

retreive 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 annotatins.
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

---

getSmilesInchi *getSmilesInchi*

Description

use PubChem API to look up full smiles and inchi notation for each inchikey

Usage

```
getSmilesInchi(ramclustObj = NULL, inchikey = NULL, ignore.stereo = TRUE)
```

Arguments

ramclustObj	ramclustR object to look up smiles and inchi for each inchikey (without a smiles/inchi). Must provide one of ramclustObj or inchikey.
inchikey	character vector of inchikey strings. Must provide one of ramclustObj or inchikey.
ignore.stereo	logical. default = TRUE. If the Pubchem databases does not have the full inchikey string, should we search by the first (non-stereo) block of the inchikey? When true, returns the first pubchem match to the inchikey block one string. If the full inchikey is present, that is used preferentially.

Details

The $inchikey slot is used to look up parameters from pubchem. PubChem CID, a pubchem URL, smiles (canonical) and inchi are returned. if smiles and inchi slots are alread present (from MS-Finder, for example) pubchem smiles and inchi are used to fill in missing values only, not replace.

Value

returns a ramclustR object. new vector of $smiles and $inchi with length equal to number of compounds.

Author(s)

Corey Broeckling

References

Kim S, Thiessen PA, Bolton EE, Bryant SH. PUG-SOAP and PUG-REST: web services for programmatic access to chemical information in PubChem. Nucleic Acids Res. 2015;43(W1):W605-11.

---

import.msfinder.formulas

import.msfinder.formulas

Description

After running MSFinder on .mat or .msp files, import the formulas that were predicted and their scores

Usage

```
import.msfinder.formulas(ramclustObj = NULL, mat.dir = NULL, msp.dir = NULL)
```

Arguments

ramclustObj	R object - the ramclustR object which was used to write the .mat or .msp files
mat.dir	optional path to .mat directory
msp.dir	optional path to .msp directory

Details

this function imports the output from the MSFinder program to support annotation of the ramclustR object

Value

new slot at $msfinder.formula.details

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.

Tsugawa H, Kind T, Nakabayashi R, Yukihira D, Tanaka W, Cajka T, Saito K, Fiehn O, Arita M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem. 2016 Aug 16;88(16):7946-58. doi: 10.1021/acs.analchem.6b00770. Epub 2016 Aug 4. PubMed PMID: 27419259.

`import.msfinder.mssearch`

import.MSFinder.mssearch

Description

After running MSFinder on .mat or .msp files, import the spectral search results

Usage

```
import.msfinder.mssearch(ramclustObj = NULL, mat.dir = NULL, msp.dir = NULL)
```

Arguments

`ramclustObj`	R object - the ramclustR object which was used to write the .mat or .msp files
`mat.dir`	optional path to .mat directory
`msp.dir`	optional path to .msp directory

Details

this function imports the output from the MSFinder program to annotate the ramclustR object

Value

an updated ramclustR object, with new slots at $msfinder.mssearch.details and $msfinder.mssearch.scores

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.

Tsugawa H, Kind T, Nakabayashi R, Yukihira D, Tanaka W, Cajka T, Saito K, Fiehn O, Arita M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem. 2016 Aug 16;88(16):7946-58. doi: 10.1021/acs.analchem.6b00770. Epub 2016 Aug 4. PubMed PMID: 27419259.

---

import.msfinder.structures

write.methods

Description

write RAMClustR processing methods and citations to text file

Usage

```
import.msfinder.structures(ramclustObj = NULL, mat.dir = NULL, msp.dir = NULL)
```

Arguments

ramclustObj	R object - the ramclustR object which was used to write the .mat or .msp files
mat.dir	directory in which to look for mat file MSFinder output - by default the /spectra/mat in the working directory
msp.dir	directory in which to look for msp file MSFinder output - by default the /spectra/msp in the working directory

Details

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

---

impRamSearch *impRamSearch*

Description

import ramsearch output for annotating an RC object

Usage

```
impRamSearch(ramclustObj = NULL, ramsearchout = "spectra/results.rse")
```

Arguments

ramclustObj	ramclustR object to annotate
ramsearchout	path to .rse file to import

Details

Annotation of ramclustR exported .msp spectra is accomplished using RAMSearch. Exported ramsearch annotations (.rse) can be imported with this function

Value

returns a ramclustR object. new slots holding .rse 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.

`manual.annotation.template`
manual.annotation.template

Description

export a .csv formatted template for manually editing MSFinder annotations

Usage

```
manual.annotation.template(
  ramclustObj = NULL,
  outfile = "manual.annotation.template.csv"
)
```

Arguments

`ramclustObj`	ramclustR object to annotate
`outfile`	output file directory and name. default = 'manual.annotation.template.csv'

Details

While unsupervised annotation is rapid and objective, subjective knowledge can be used to improve annotations. This function writes a template file containing compound name, computationally assigned inchikey, and an empty column for your manually inferred inchikey. Upon completion of manual annotation, you can reimport this file and update your ramclustR object to reflect your manual input.

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.

Tsugawa H, Kind T, Nakabayashi R, Yukihira D, Tanaka W, Cajka T, Saito K, Fiehn O, Arita M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem. 2016 Aug 16;88(16):7946-58. doi: 10.1021/acs.analchem.6b00770. Epub 2016 Aug 4. PubMed PMID: 27419259.

`mergeRCobjects` *mergeRCobjects*

Description

merge two ramclustR objects

Usage

```
mergeRCobjects(
  ramclustObj.1 = NULL,
  ramclustObj.2 = NULL,
  mztol = 0.2,
  rttol = 30,
  course.rt.adj = NULL,
  mzwt = 2,
  rtwt = 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 racmlustObj1. That is only features which appear consistent with those from ramclustObj.1 will be retained.
`mztol`	numeric: absolute mass tolerance around mz
`rttol`	numeric: feaure 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 ramlcustObj 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

`ramclustR` *ramclustR*

Description

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

Usage

```
ramclustR(
  xcmsObj = NULL,
  ms = NULL,
  idmsms = 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,
  fftempdir = NULL,
  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
idmsms	filepath: optional idMSMS / MSe csv data. same dim and names as ms required
taglocation	character: "filepaths" by default, "phenoData[,1]" is another option. referse 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 agressively 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: tranfer phenotype data from XCMS object to SpecAbund dataset?
mspout	logical: write msp formatted specta 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 defualt 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.
fftempdir	valid path: if there are file size limitations on the default ff pacakge temp directory - getOptions('fftempdir') - you can change the directory used as the fftempdir with this option.
replace.zeros	logincal: 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

$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

$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 witht he 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.

`rc.calibrate.ri` *rc.cmpd.filter.cv*

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 columnn names (case sensitive)
`poly.order`	integer default = 3. polynomical order used to fit rt vs ri data, and calculate ri for all feature and metabolite rt values. poly.order should be apprciably 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 ($fri) and compounds ($clri).

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 = 0.3)
```

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 maximumn 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.get.classyfire` *getClassyFire*

Description

use classyfire web API to look up full ClassyFire heirarchy for each inchikey

Usage

```
rc.cmpd.get.classyfire(
  ramclustObj = NULL,
  inchikey = NULL,
  get.all = TRUE,
  max.wait = 10,
  posts.per.minute = 5
)
```

Arguments

`ramclustObj`	ramclustR object to ClassyFy. Must supply one of either ramclustObj or inchikey (see below)
`inchikey`	vector of text inchikeys to ClassyFy. Must supply one of either ramclustObj or inchikey.
`get.all`	logical; if TRUE, when inchikey classyfire lookup fails, submits for classyfication. Can be slow. max.wait (below) sets max time to spend on each compound before moving on. default = FALSE.
`max.wait`	numeric; maximum time (seconds) to wait per compound when 'get.all' = TRUE.
`posts.per.minute`	integer; a limit set when 'get.all' is true. ClassyFire server accepts no more than 5 posts per minute when calculating new ClassyFire results. Slows down submission process to keep server from denying access.

Details

The $inchikey slot is used to look up the

Value

returns a ramclustR object. new dataframe in $classyfire slot with rows equal to number of compounds.

Author(s)

Corey Broeckling

References

Djoumbou Feunang Y, Eisner R, Knox C, Chepelev L, Hastings J, Owen G, Fahy E, Steinbeck C, Subramanian S, Bolton E, Greiner R, and Wishart DS. ClassyFire: Automated Chemical Classification With A Comprehensive, Computable Taxonomy. Journal of Cheminformatics, 2016, 8:61. DOI: 10.1186/s13321-016-0174-y

---

`rc.cmpd.get.pubchem` *rc.cmpd.get.pubchem*

Description

use pubchem rest and view APIs to retreive structures, CIDs (if a name or inchikey is given), synonyms, and optionally vendor data, when available.

Usage

```
rc.cmpd.get.pubchem(
  ramclustObj = NULL,
  search.name = NULL,
  cmpd.names = NULL,
  cmpd.cid = NULL,
  cmpd.inchikey = NULL,
  use.parent.cid = FALSE,
  manual.entry = FALSE,
  get.vendors = FALSE,
 priority.vendors = c("Sigma Aldrich", "Alfa Chemistry", "Acros Organics", "VWR",
    "Alfa Aesar", "molport", "Key Organics", "BLD Pharm"),
  get.properties = TRUE,
  all.props = TRUE,
  get.synonyms = TRUE,
  find.short.lipid.name = TRUE,
  find.short.synonym = TRUE,
  max.name.length = 30,
  assign.short.name = TRUE,
  get.bioassays = FALSE,
  write.csv = TRUE
)
```

Arguments

`ramclustObj`	RAMClust Object input. if used, ramclustObj$CID, ramclustObj$inchikey, and ramclustObj$ann are used as input, in that order, and ramclustObj is returned with $pubchem slot appended.
`search.name`	character. optional name to assign to pubchem search to name output .csv files.
`cmpd.names`	character vector. i.e. c("caffeine", "theobromine", "glucose")
`cmpd.cid`	numeric integer vector. i.e. c(2519, 5429, 107526)

`cmpd.inchikey` character vector. i.e. c("RYYVLZVUVIJVGH-UHFFFAOYSA-N", "YAPQBXQYLJRXSA-UHFFFAOYSA-N", "GZCGUPFRVQAUEE-SLPGGIOYSA-N")

`use.parent.cid` logical. If TRUE, the CID for each supplied name/inchikey is used to retreive its parent CID (i.e. the parent of sodium palmitate is palmitic acid). The parent CID is used to retrieve all other names, properties.

`manual.entry` logical. if TRUE, user input is enabled for compounds not matched by name. A browser window will open with the pubchem search results in your default browser.

`get.vendors` logical. if TRUE, vendor data is returned for each compound with a matched CID. Includes vendor count and vendor product URL, if available

`priority.vendors` charachter vector. i.e. c("MyFavoriteCompany", "MySecondFavoriteCompany"). If these vendors are found, the URL returned is from priority vendors. Priority is given by order input by user.

`get.properties` logical. if TRUE, physicochemical property data are returned for each compound with a matched CID.

`all.props` logical. If TRUE, all pubchem properties (https://pubchemdocs.ncbi.nlm.nih.gov/pug-rest$_Toc494865567) are returned. If false, only a subset (faster).

`get.synonyms` = TRUE. logical. if TRUE, retrieve pubchem synonyms. returned to $synonyms slot

`find.short.lipid.name` = TRUE. logical. If TRUE, and get.synonyms = TRUE, looks for lipid short hand names in synonymns list (i.e. PC(36:6)). returned to $short.name slot. Short names are assigned only if assign.short.names = TRUE.

`find.short.synonym` = TRUE. logical. If TRUE, and get.synonyms = TRUE, looks for lipid short synonymns, with prioritization for names with fewer numeric characters (i.e. database accession numbers or CAS numbers). returned to $short.name slot. Short names are assigned only if assign.short.names = TRUE.

`max.name.length` = 20. integer. If names are longer than this value, short names will be searched for, else, retain original name.

`assign.short.name` = TRUE. If TRUE, short names from find.short.lipid.name and/or find.short.synonym = TRUE, short names are assigned the be the default annotation name ($ann slot), and original annotations are moved to $long.name slot.

`get.bioassays` logical. If TRUE, return a table summarizing existing bioassay data for that CID.

`write.csv` logical. If TRUE, write csv files of all returned pubchem data.

Details

useful for moving from chemical name to digital structure represtation. greek letters are assumed to be 'UTF-8' encoded, and are converted to latin text before searching. if you are reading in your compound name list, do so with 'encoding' set to 'UTF-8'.

Value

returns a list with one or more of $puchem (compound name and identifiers) - one row in dataframe per CID; $properties contains pysicochemical properties - one row in dataframe per CID; $vendors contains the number of vendors for a given compound and selects a vendor based on 'priortity.vendors' supplied, or randomly choses a vendor with a HTML link - one row in dataframe per CID; $bioassays contains a summary of bioassay activity data from pubchem - zero to many rows in dataframe per CID

Author(s)

Corey Broeckling

---

rc.cmpd.get.smiles.inchi

getSmilesInchi

Description

use PubChem API to look up full smiles and inchi notation for each inchikey

Usage

```
rc.cmpd.get.smiles.inchi(
  ramclustObj = NULL,
  inchikey = NULL,
  ignore.stereo = TRUE
)
```

Arguments

ramclustObj	ramclustR object to look up smiles and inchi for each inchikey (without a smiles/inchi). Must provide one of ramclustObj or inchikey.
inchikey	character vector of inchikey strings. Must provide one of ramclustObj or inchikey.
ignore.stereo	logical. default = TRUE. If the Pubchem databases does not have the full inchikey string, should we search by the first (non-stereo) block of the inchikey? When true, returns the first pubchem match to the inchikey block one string. If the full inchikey is present, that is used preferentially.

Details

The $inchikey slot is used to look up parameters from pubchem. PubChem CID, a pubchem URL, smiles (canonical) and inchi are returned. if smiles and inchi slots are alread present (from MS-Finder, for example) pubchem smiles and inchi are used to fill in missing values only, not replace.

Value

returns a ramclustR object. new vector of $smiles and $inchi with length equal to number of compounds.

Author(s)

Corey Broeckling

References

Kim S, Thiessen PA, Bolton EE, Bryant SH. PUG-SOAP and PUG-REST: web services for programmatic access to chemical information in PubChem. Nucleic Acids Res. 2015;43(W1):W605-11.

`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)
```

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 on in console. If character vector (i.e. c("trt", "time")) names are assigned to table

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 invididual 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

extractor for xcms objects in preparation for clustering

Usage

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

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 maximumn 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.

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.

ramclustR object with total extracted ion normalized data.

Author(s)

Corey Broeckling

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, 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.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.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
)
```

Arguments

`ramclustObj`	ramclustObj containing MSdata with optional MSMSdata (MSe, DIA, idMSMS)
`replace.int`	default = 0.2. proportion of minimum feature value to replace NA (or zero) values with
`replace.noise`	default = 0.2. proportion ofreplace.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

Details

noise is added by finding for each feature the mimimum 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.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. referse 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 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 (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 be 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.qc` *rc.qc*

Description

summarize quality control for clustering and for quality control sample varation 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 calcuate 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 diaganol indicate either poor clustering or a group of coregulated metabolites with similar retention time. It is an imperfect diagnostic, partuclarly 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 int he 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` *rc.ramclustr*

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,
  fftempdir = NULL
)
```

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 agressively 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 defualt 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.

`fftempdir` valid path: if there are file size limitations on the default ff pacakge temp directory - getOptions('fftempdir') - you can change the directory used as the fftempdir with this option.

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 varation 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.qamples*

Description

summarize quality control for clustering and for quality control sample varation 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 varation

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 calcuate 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 diaganol indicate either poor clustering or a group of coregulated metabolites with similar retention time. It is an imperfect diagnostic, partuclarly 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 int he 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.

`write.methods` *write.methods*

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 relevent 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

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.

Details

exports files to a directory called 'spectra'. If one.file = FALSE, a new directory 'spectra/msp' is created to hold the invididual 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

Index