Supplementary Information

Package 'LargeMetabo'

Type Package

Version 0.1.0

Title an out-of-the-box tool for processing and analyzing large-scale metabolomic data

Author Qingxia Yang

Maintainer Qingxia Yang <yangqx@cqu.edu.cn>

Depends R (>= 3.0), dynamicTreeCut (>= 1.62), fastcluster

Imports AUC, genefilter, d3heatmap, ggplot2, ggfortify, factoextra, igraph, FSelector, varSelRF, corrplot, mixOmics, ropls, CluMSID, siggenes, e1071, SOMbrero, sampling

Suggests rmarkdown, knitr, hexbin, shinycssloaders, shinyBS, shinyjs, shinyjqui

Description This package for processing and analyzing large-scale metabolomic data, including data combination for metabolomics data, biomarker selection, enhanced metabolite annotation and enrichment analysis.

License MIT + file LICENSE

URL https://github.com/LargeMetabo/LargeMetabo

RoxygenNote 7.1.2

NeedsCompilation no

LazyData true

R topics documented:

Integrate_Data 2
Removal_Batch 3
Sample_Separation 4
Marker_Identify 5
Marker_Assess 5
MetaboAnnotation 6
AnnotaTandem 7
annotaDataTandem 8
AnnotaTandem_plot 9
KEGGEnrichPlotPanel 10
Enrichment 11
KEGGEnrichPlot 12
EnrichPlot 13

Index

Integrate_Data	Integrating multiple datasets

Description

Integrating multiple datasets from different analytical experiments

Usage

Integrate_Data(mutile_Group, RT_Tolerance_1 = 10, mz_Tolerance_1 = 0.1, RT_Tolerance_2 = 10, mz_Tolerance_2 = 0.1)

Arguments

mutile_Group	Multiple datasets from multiple groups for integrating as a comprehensive dataset.
RT_Tolerance_1	Set the tolerance of retention time for data integration in the primary phase.
mz_Tolerance_1	Set the tolerance of mass-to-charge ratio for data integration in the primary phase.

RT_Tolerance_2 Set the tolerance of retention time for data integration in the secondary phase.

mz_Tolerance_2 Set the tolerance of mass-to-charge ratio for data integration in the secondary phase.

Value

A matrix of a comprehensive dataset integrated form multiple groups is returned.

Author(s)

Qingxia Yang <yangqx@cqu.edu.cn>

Examples

Integrate_Data(mutile_Group, RT_Tolerance_1 = 10, mz_Tolerance_1 = 0.1, RT_Tolerance_2 = 10, mz_Tolerance_2 = 0.1)

Removal_Batch	Remove batch effects for the integrated dataset

---

Description

Remove batch effects among different analytical experiments for the integrated dataset

Usage

Removal_Batch(mutile_align, n = 3, algorithm = "BMC/PAMR")

Arguments

mutile_align The comprehensive dataset integrated from multiple groups.

n The number of multiple groups.

algorithm The algorithm for removing batch effects in multiple groups. The algorithm could be BMC/PAMR, ComBat/EB, GlobalNorm, Log2 or None.

Value

A matrix of a dataset after removing batch effects for the integrated dataset.

Author(s)

Qingxia Yang <yangqx@cqu.edu.cn>

Examples

Removal_Batch(mutile_align, n = 3, algorithm = "BMC/PAMR")

Sample_Separation	Plot of sample separation

Description

Plot of sample separation for visualizing the clustering and separation of different samples

Usage

Sample_Separation(finalData, finalLabel, clusters=2, method = "HCA")

Arguments

finalData	The matrix of dataset for sample separation.
finalLabel	The label of dataset for sample separation.
clusters	The number of cluster for sample separation.
method	The method for sample separation. The method could be HCA, KMC, PCA and SOM.

Value

The plot of sample separation.

Author(s)

Qingxia Yang <yangqx@cqu.edu.cn>

Examples

finalData <- MarkerData$finalData

finalLabel <- MarkerData$finalLabel

Sample_Separation(finalData, finalLabel, clusters=2, method = "HCA")

Marker_Identify	Marker identification in large-scale metabolomics data

Description

Marker identification in large-scale metabolomics data

Usage

Marker_Identify(finalData, finalLabel, method = "FC")

Arguments

finalData	The matrix of dataset for marker identification.
finalLabel	The label of dataset for marker identification.
method	The method for marker identification. The method could be FC, PLS-DA, OPLS-DA, t-test, CHIS, CFS, ENTROPY, LMEB, RELIEF, RF, SAM, SVMRFE or WRST.

Value

The result of marker identification for metabolomics data.

Author(s)

Qingxia Yang <yangqx@cqu.edu.cn>

Examples

finalData <- MarkerData$finalData

finalLabel <- MarkerData$finalLabel

Marker_Identify(finalData, finalLabel, method = "FC")

Marker_Assess	Assessing the markers of different methods by classification

Description

Classification assessment for the markers identified using different marker identification methods in large-scale metabolomics data

Usage

Marker_Identify(finalData, finalLabel, method = "PLS-DA")

Arguments

finalData	The matrix of dataset for marker identification.
finalLabel	The label of dataset for marker identification.
method	The method for marker identification. The method could be FC, PLS-DA, OPLS-DA, t-test, CHIS, CFS, ENTROPY, LMEB, RELIEF, RF, SAM, SVMRFE or WRST.

Value

The ROC curve and AUC value in the SVM classification using markers.

Author(s)

Qingxia Yang <yangqx@cqu.edu.cn>

Examples

finalData <- MarkerData$finalData

finalLabel <- MarkerData$finalLabel

Marker_Assess(finalData, finalLabel, method = "PLS-DA")

MetaboAnnotation Metabolite annotation for MS1 spectra

Description

Metabolite annotation for MS1 spectra based on a metabolite database

Usage

MetaboAnnotation(AnnotaData, masstole = 10, toleUnit = 1, annotaDB = "metlin", ionMode = "pos")

Arguments

AnnotaData	The value of mass-to-charge ratio for metabolite annotation.
masstole	The tolerance of m/z for metabolite annotation.

toleUnit	The unit of m/z for metabolite annotation. The unit could be 1 for Da, or 2 for ppm.
annotaDB	The database of m/z for metabolite annotation. The database could be metlin or hmdb.
ionMode	The mode of m/z for annotation. The mode could be pos, neg or neu.

Value

The list of results of metabolite annotation for different charges for primary mass spectrometry.

Author(s)

Qingxia Yang <yangqx@cqu.edu.cn>

Examples

```
AnnotaMS <- AnnotaData$AnnotaMS

MetaboA_res <- MetaboAnnotation(AnnotaMS)

MetaboA_res$`M+H-2H2O`
```

AnnotaTandem Parameters of metabolite annotation for MS/MS

Description

Parameters of metabolite annotation of tandem mass spectrum (MS/MS)

Usage

```
AnnotaTandem(Parent_mass = 181.04, TandemData, massTandem = 0.1, toleUnitTandem = 1, massmzTandem = 0.5, toleUnitmzTandem = 1, ModeTandem = "Positive", ionEnergy = "low(10V)")
```

Arguments

Parent_mass	The value of parent ion mass for metabolite annotation of tandem mass spectrum.
TandemData	The value of MS/MS peak list (m/z & Intensity) for metabolite annotation of tandem mass spectrum.

massTandem	The tolerance of parent ion mass for metabolite annotation of tandem mass spectrum.
toleUnitTandem	The unit of parent ion mass for metabolite annotation of tandem mass spectrum. The unit could be 1 for Da, or 2 for ppm.
massmzTandem	The tolerance of MS/MS peak mass for metabolite annotation of tandem mass spectrum.
toleUnitmzTandem	The unit of MS/MS peak mass for metabolite annotation of tandem mass spectrum. The unit could be 1 for Da, or 2 for ppm.
ModeTandem	The ionization mode for metabolite annotation of tandem mass spectrum. The mode could be Positive or Negative.
ionEnergy	The CID energy for metabolite annotation of tandom mass spectrum. The mode could be low(10V) Medium(25V) High(40V) or All.

Value

The parameters of metabolite annotation of tandem mass spectrum (MS/MS).

Author(s)

Qingxia Yang <yangqx@cqu.edu.cn>

Examples

```
Parent_mass <- AnnotaData$Parent_mass

TandemData <- AnnotaData$TandomData

AnnotaParamTandem <- AnnotaTandem(Parent_mass, TandemData)
```

---

annotaDataTandem Table of metabolite annotation for MS/MS

Description

The table of results for metabolite annotation of tandem mass spectrum (MS/MS).

Usage

```
annotaDataTandem(AnnotaParamTandem)
```

Arguments

AnnotaParamTandem The parameters of metabolite annotation of tandem mass spectrum.

Value

The table of results for metabolite annotation of tandem mass spectrum.

Author(s)

Qingxia Yang <yangqx@cqu.edu.cn>

Examples

```
Parent_mass <- AnnotaData$Parent_mass

TandemData <- AnnotaData$TandomData

AnnotaParamTandem <- AnnotaTandem(Parent_mass, TandemData)

annotaDataTandem(AnnotaParamTandem)
```

AnnotaTandem_plot Plot of metabolite annotation of MS/MS

Description

The plot of results for metabolite annotation of tandem mass spectrum (MS/MS).

Usage

```
AnnotaTandem_plot(AnnotaParamTandem, TandemData)
```

Arguments

AnnotaParamTandem The parameters of metabolite annotation of tandem mass spectrum.

TandemData The value of MS/MS peak list (m/z & Intensity) for metabolite annotation of tandem mass spectrum.

Value

The plot of results for metabolite annotation of tandem mass spectrum.

Author(s)

Qingxia Yang <yangqx@cqu.edu.cn>

Examples

```
Parent_mass <- AnnotaData$Parent_mass

TandemData <- AnnotaData$TandomData

AnnotaParamTandem <- AnnotaTandem(Parent_mass, TandemData)

AnnotaTandem_plot(AnnotaParamTandem, TandemData)
```

KEGGEnrichPlotPanel Parameters for metabolite enrichment

Description

The parameters for metabolite enrichment based on the metabolite database.

Usage

```
KEGGEnrichPlotPanel(sampleData, enrichDB = "kegg", pvalcutoff = 0.05, IDtype = 1, cateIdx = 1)
```

Arguments

sampleData	The charactor of input for metabolite enrichment.
enrichDB	The database of the input data. The database could be kegg, smpdb, cfam, foodb, biofunc, tcm, spectax or toxin.
pvalcutoff	The cutoff of p value for metabolite enrichment.
IDtype	The number for name type (1, 2, 3…) for metabolite enrichment, such as KEGG ID, CAS ID, PubChem ID, Name or HMDB ID. If enrichDB is 'kegg', IDtype is 1 (KEGG ID), 2 (CAS ID), 3 (PubChem ID), or 4 (Compound Name). If enrichDB is 'smpdb', IDtype is 1 (HMDB ID), 2 (KEGG ID), 3 (CAS ID), 4 (PubChem ID), or 5 (Compound Name). If enrichDB is 'cfam', IDtype is 1 (CFAM ID), 2 (CAS ID), 3 (PubChem ID), 4 (Compound Name), or 5 (HMDB ID). If enrichDB is 'foodb', IDtype is 1 (FoodDB ID), 2 (CAS ID), 3 (PubChem ID), 4 (Compound Name) or 5 (HMDB ID). If enrichDB is 'biofunc', IDtype is 1 (HMDB ID), 2 (KEGG ID), 3 (CAS ID), 4 (PubChem ID) or 5 (Compound Name). If enrichDB is 'tcm', IDtype is 1 (LargeMetabo ID), 2 (PubChem ID), 3(Compound

Name), or 4 (CAS ID). If enrichDB is 'spectax', IDtype is 1 (LargeMetabo ID), 2 (PubChem ID), 3(Compound Name), 4 (CAS ID) or 5 (HMDB ID). If enrichDB is 'toxin', IDtype is 1 (T3DB ID), 2 (PubChem ID), 3 (Compound Name), 4 (CAS ID) 5 (HMDB ID) or 6 (KEGG ID).

cateIdx The number of category for metabolite enrichment.

Value

The parameters for metabolite enrichment.

Author(s)

Qingxia Yang <yangqx@cqu.edu.cn>

Examples

sampleDatakegg <- EnrichData$sampleDatakegg

KEGGEnrichPlotPanel(sampleDatakegg, enrichDB = "kegg", pvalcutoff = 0.05, IDtype = 1, cateIdx = 1)

Enrichment Table for metabolite enrichment

Description

The table for metabolite enrichment.

Usage

Enrichment(param)

Arguments

param The parameters for metabolite enrichment.

Value

The table of results for metabolite enrichment.

Author(s)

Qingxia Yang <yangqx@cqu.edu.cn>

Examples

sampleDatakegg <- EnrichData$sampleDatakegg

EnrichParam <- KEGGEnrichPlotPanel(sampleDatakegg, enrichDB = "kegg", pvalcutoff = 0.05, IDtype = 1, cateIdx = 1)

EnrichResultList <- Enrichment(EnrichParam); EnrichResultList$Table.Result

---

KEGGEnrichPlot	Plot of metabolite enrichment using KEGG database

---

Description

The plot of metabolite enrichment using KEGG database.

Usage

KEGGEnrichPlot(EnrichResultList, cpdID, cpdFC)

Arguments

EnrichResultList	The table of results for metabolite enrichment using KEGG database.
cpdID	The character of input for metabolite enrichment.
cpdFC	The distance in the metabolite enrichment plot using KEGG database.

Value

The plot of results for metabolite enrichment using KEGG database.

Author(s)

Qingxia Yang <yangqx@cqu.edu.cn>

Examples

sampleDatakegg <- EnrichData$sampleDatakegg

EnrichParam <- KEGGEnrichPlotPanel(sampleDatakegg, enrichDB = "kegg", pvalcutoff = 0.05, IDtype = 1, cateIdx = 1)

EnrichResultList <- Enrichment(EnrichParam)

EnrichFC <- seq(from = -2, to = 2, length.out = 24)

KEGGEnrichPlot(EnrichResultList = EnrichResultList, cpdID = sampleDatakegg, cpdFC = EnrichFC)

EnrichPlot	Plot of results for metabolite enrichment

Description

The plot of results for metabolite enrichment.

Usage

EnrichPlot(db, EnrichResultList)

Arguments

db	The database of input for metabolite enrichment.
EnrichResultList	The table of results for metabolite enrichment.

Value

The plot of results for metabolite enrichment.

Author(s)

Qingxia Yang <yangqx@cqu.edu.cn>

Examples

sampleDatacas <- EnrichData$sampleDatacas

enrichDB <- EnrichData$enrichDB

EnrichParam <- KEGGEnrichPlotPanel(sampleDatacas, enrichDB = enrichDB, pvalcutoff = 0.05, IDtype = 2, cateIdx = 1)

EnrichResultList <- Enrichment(EnrichParam)

dbChoice <- enrichDB

EnrichPlot(dbChoice, EnrichResultList)