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@njupt.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, magrittr, dplyr, readr

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

Description This package for processing and analyzing large-scale metabolomic data including data integration for multiple analytical experiments, performance assessment of biomarker identification, 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:

Index

Integrate_Data	Data integration for multiple analytical experiments

---

Description

Integrating multiple datasets from different analytical experiments

Usage

Integrate_Data(MutileGroup, RTTolerance1 = 10, mzTolerance1 = 0.1, RTTolerance2 = 10, mzTolerance2 = 0.1)

Arguments

MutileGroup	Data list including multiple datasets from multiple experiments and are integrated as a comprehensive dataset.
RTTolerance1	Set the tolerance of retention time in the first step of data integration.
mzTolerance1	Set the tolerance of mass-to-charge ratio in the first step of data integration.
RTTolerance2	Set the tolerance of retention time in the second step of data integration.
mzTolerance2	Set the tolerance of mass-to-charge ratio in the second step of integration.

Value

A matrix of an integrated dataset form multiple analytical experiments is returned.

Author(s)

Qingxia Yang <yangqx@njupt.edu.cn>

Examples

```
AlignData <- Integrate_Data(MutileGroup, RTTolerance1 = 10, mzTolerance1 = 0.1,
RTTolerance2 = 10, mzTolerance2 = 0.1)

head(AlignData)
```

Removal_Batch	Remove batch effects for the integrated dataset

---

Description

Remove batch effects among different analytical experiments for the integrated dataset

Usage

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

Arguments

MutileAlign	The integrated dataset from multiple analytical experiments.
n	The number of multiple analytical experiments.
algorithm	The algorithm for removing batch effects in multiple analytical experiments. The algorithm can be BMC/PAMR, ComBat/EB, GlobalNorm or None.

Value

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

Author(s)

Qingxia Yang <yangqx@njupt.edu.cn>

Examples

```
DataAfterBatch <- Removal_Batch(MutileAlign, n = 3, algorithm = "BMC/PAMR")

head(DataAfterBatch)
```

Sample_Separation	Plot of sample separation for large-scale metabolomic data

Description

Plot of sample separation to visualize the clustering and separation of all 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 can be HCA, KMC, PCA or SOM.

Value

The plot of sample separation.

Author(s)

Qingxia Yang <yangqx@njupt.edu.cn>

Examples

```
finalData <- MarkerData$finalData

finalLabel <- MarkerData$finalLabel

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

Marker_Identify	Marker identification for large-scale metabolomic data

Description

Marker identification for large-scale metabolomic 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 can 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 large-scale metabolomic data.

Author(s)

Qingxia Yang <yangqx@njupt.edu.cn>

Examples

```
finalData <- MarkerData$finalData

finalLabel <- MarkerData$finalLabel

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

Marker_Assess Performance assessment for marker identification methods

---

Description

Performance assessment for the markers identified by marker identification method in large-scale metabolomics data based on the SVM classification model

Usage

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

Arguments

finalData	The matrix of dataset for performance assessment of marker identification.
finalLabel	The label of dataset for performance assessment of marker identification.
method	The method used for identifying markers for performance assessment. The method can 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 model using identified markers.

Author(s)

Qingxia Yang <yangqx@njupt.edu.cn>

Examples

```
finalData <- MarkerData$finalData

finalLabel <- MarkerData$finalLabel

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

Metabo_Annotation	Metabolite annotation for MS1 spectra

Description

Metabolite annotation for MS1 spectra based on a new metabolite database

Usage

Metabo_Annotation(AnnotaMS, masstole = 10, toleUnit = 1, annotaDB = "metlin", ionMode = "pos")

Arguments

AnnotaMS	The value of mass-to-charge ratio for MS1 spectra for metabolite annotation.
masstole	The tolerance of m/z for metabolite annotation, and the unit is ppm or Da.
toleUnit	The unit of m/z for metabolite annotation. The unit is 1 for Da and 2 for ppm.
annotaDB	The database for metabolite annotation. The database can be metlin or hmdb.
ionMode	The mode of m/z for metabolite annotation. The mode can be pos, neg or neu.

Value

The result of metabolite annotation for the adducts and charges for MS1 spectra.

Author(s)

Qingxia Yang <yangqx@njupt.edu.cn>

Examples

```
AnnotaMS <- AnnotaData$AnnotaMS

MetaboAResult <- Metabo_Annotation(AnnotaMS)

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

Annota_Tandem	Parameters of metabolite annotation for MS/MS

Description

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

Usage

Annota_Tandem(ParentMass, TandemData, massTandem = 0.1, toleUnitTandem = 1, massmzTandem = 0.5, toleUnitmzTandem = 1, ModeTandem = "Positive", ionEnergy = "low(10V)")

Arguments

ParentMass	The m/z value of parent ion for metabolite annotation of MS/MS.
TandemData	The MS/MS peak list (including the matrix of m/z values and intensities of MS/MS) for metabolite annotation of MS/MS.
massTandem	The tolerance of parent ion for metabolite annotation of MS/MS, and the unit is ppm or Da.
toleUnitTandem	The unit of parent ion for metabolite annotation of MS/MS. The unit is 1 for Da and 2 for ppm.
massmzTandem	The tolerance of MS/MS peak for metabolite annotation of MS/MS, and the unit is ppm or Da.
toleUnitmzTandem	The unit of MS/MS peak for metabolite annotation of MS/MS. The unit is 1 for Da and 2 for ppm.
ModeTandem	The ionization mode for metabolite annotation of tandem mass spectrum. The mode can be Positive or Negative.
ionEnergy	The CID energy for metabolite annotation of MS/MS. The mode can be low(10V), Medium(25V), High(40V), or All.

Value

The parameters for metabolite annotation of MS/MS.

Author(s)

Qingxia Yang <yangqx@njupt.edu.cn>

Examples

```
ParentMass <- AnnotaData$ParentMass

TandemData <- AnnotaData$TandomData

AnnotaParamTandem <- Annota_Tandem(ParentMass, TandemData)
```

---

annota_Data_Tandem Table of metabolite annotation for MS/MS

Description

The table of metabolite annotation of MS/MS.

Usage

annota_Data_Tandem(AnnotaParamTandem)

Arguments

AnnotaParamTandem The parameters of metabolite annotation of MS/MS.

Value

The table of metabolite annotation of MS/MS.

Author(s)

Qingxia Yang <yangqx@njupt.edu.cn>

Examples

```
ParentMass <- AnnotaData$ParentMass

TandemData <- AnnotaData$TandomData

AnnotaParamTandem <- Annota_Tandem(ParentMass, TandemData)

annota_Data_Tandem(AnnotaParamTandem)
```

Annota_Tandem_plot	Plot of metabolite annotation of MS/MS

Description

The plot of results for metabolite annotation of MS/MS.

Usage

Annota_Tandem_plot(AnnotaParamTandem, TandemData)

Arguments

AnnotaParamTandem	The parameters of metabolite annotation of MS/MS.
TandemData	The MS/MS peak list (including the matrix of m/z values and intensities of MS/MS) for metabolite annotation of MS/MS.

Value

The plot of results for metabolite annotation of MS/MS.

Author(s)

Qingxia Yang <yangqx@njupt.edu.cn>

Examples

```
ParentMass <- AnnotaData$ParentMass

TandemData <- AnnotaData$TandomData

AnnotaParamTandem <- Annota_Tandem(ParentMass, TandemData)

Annota_Tandem_plot(AnnotaParamTandem, TandemData)
```

KEGG_Enrich_PlotPanel	Parameters for enrichment analysis by the metabolite database

Description

The parameters for enrichment analysis based on the metabolite database.

Usage

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

Arguments

sampleDatakegg	The character of input data for enrichment analysis.
enrichDB	The database for enrichment analysis. The database can be kegg, smpdb, cfam, foodb, biofunc, tcm, spectax or toxin.
pvalcutoff	The cutoff of *p* value for enrichment analysis.
IDtype	The number of type for names (1, 2, 3…) for enrichment analysis. 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 (1, 2, 3…) of category for metabolite enrichment. If enrichDB is 'cfam', cateIdx is 1 (Class), 2 (Superfamily), or 3 (Family). If enrichDB is 'foodb', cateIdx is 1 (Food group), or 2 (Food subgroup). If enrichDB is 'biofunc', cateIdx is 1 (Blood), 2 (Urine), or 3 (Cerebrospinal). If enrichDB is 'tcm', cateIdx is 1 (TCM primary category), or 2 (TCM secondary category). If enrichDB is 'spectax', cateIdx is 1 (Superkingdom), 2 (Kingdom), 3(Phylum), 4 (Class), 5 (Order), 6 (Family), 7 (Genus), or 8 (Species).

Value

The parameters for enrichment analysis.

Author(s)

Qingxia Yang <yangqx@njupt.edu.cn>

Examples

```
sampleDatakegg <- EnrichData$sampleDatakegg

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

Enrichment Table for enrichment analysis

Description

The table for enrichment analysis.

Usage

Enrichment(EnrichParam)

Arguments

EnrichParam The parameters for enrichment analysis.

Value

The table of results for enrichment analysis.

Author(s)

Qingxia Yang <yangqx@njupt.edu.cn>

Examples

```
sampleDatakegg <- EnrichData$sampleDatakegg

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

EnrichResultList <- Enrichment(EnrichParam)

EnrichResultList$Table.Result
```

KEGG_Enrich_Plot Plot of enrichment analysis for KEGG pathway

---

Description

The plot of enrichment analysis for KEGG pathway.

Usage

KEGG_Enrich_Plot(EnrichResultList, cpdID, cpdFC)

Arguments

EnrichResultList	The table of enrichment analysis for KEGG pathway.
cpdID	The character of input data for enrichment analysis.
cpdFC	The distance for the plot of enrichment analysis for KEGG pathway.

Value

The plot of enrichment analysis for KEGG pathway.

Author(s)

Qingxia Yang <yangqx@njupt.edu.cn>

Examples

```
sampleDatakegg <- EnrichData$sampleDatakegg

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

EnrichResultList <- Enrichment(EnrichParam)

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

KEGG_Enrich_Plot(EnrichResultList = EnrichResultList, cpdID = sampleDatakegg,
cpdFC = EnrichFC)
```

Enrich_Plot	Plot of enrichment analysis for other seven categories

---

Description

The plot of enrichment analysis for other seven categories.

Usage

Enrich_Plot(db, EnrichResultList)

Arguments

db	The database of input data for enrichment analysis for other seven categories.

EnrichResultList The table of enrichment analysis for other seven categories.

Value

The plot of enrichment analysis for other seven categories.

Author(s)

Qingxia Yang <yangqx@njupt.edu.cn>

Examples

```
sampleDatacas <- EnrichData$sampleDatacas

enrichDB <- EnrichData$enrichDB

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

EnrichResultList <- Enrichment(EnrichParam)

dbChoice <- enrichDB

Enrich_Plot(dbChoice, EnrichResultList)
```