

Package ‘AMARETTO’

November 5, 2018

Type Package

Title Multi-Omics Data Fusion for Cancer Module Discovery

Version 0.99.1

Date 2016-06-06

Author Magali Champion, Katie Planey and Olivier Gevaert

Maintainer Olivier Gevaert <olivier.gevaert@gmail.com>

Depends R (>= 3.5), RCurl, limma, foreach, parallel, doParallel, glmnet, matrixStats, RColorBrewer, impute, Matrix, BiocStyle, stringr, ComplexHeatmap, circlize, R.utils, randomcoloR, curatedTCGAData, tidyverse, callr, Rcpp, DT, htmltools, reshape2, rmarkdown

Description Module network discovery using gene expression, CNV, and methylation data.

License MIT + file LICENSE

LazyLoad yes

Encoding UTF-8

biocViews

StatisticalMethod, DifferentialMethylation, GeneRegulation, GeneExpression, MethylationArray, Transcription, Preprocessing

Suggests testthat, MASS, knitr, rmarkdown

NeedsCompilation yes

Imports callr (>= 3.0.0.9001), Rcpp

Roxygen list(markdown = TRUE)

RoxygenNote 6.1.0

LinkingTo Rcpp

VignetteBuilder knitr

Remotes r-lib/callr

R topics documented:

AMARETTO_Download	2
AMARETTO_EvaluateTestSet	2
AMARETTO_ExportResults	3
AMARETTO_HTMLreport	4
AMARETTO_Initialize	5

AMARETTO_Preprocess	6
AMARETTO_Run	7
AMARETTO_VisualizeModule	7
BatchData	8
CreateRegulatorData	8
Driver_Genes	9
MethylStates	10
ProcessedDataLIHC	10
Index	11

AMARETTO_Download	<i>AMARETTO_Download</i>
-------------------	--------------------------

Description

Downloading TCGA dataset for AMARETTO analysis

Usage

AMARETTO_Download(CancerSite, TargetDirectory, downloadData = TRUE)

Arguments

CancerSite TCGA cancer code for data download
TargetDirectory Directory path to download data

downloadData

Examples

```
TargetDirectory <- "../Downloads/" # path to data download directory  
CancerSite <- "LIHC"  
DataSetDirectories <- AMARETTO_Download(CancerSite,TargetDirectory)
```

AMARETTO_EvaluateTestSet	<i>AMARETTO_EvaluateTestSet</i>
--------------------------	---------------------------------

Description

Code to evaluate AMARETTO on a new gene expression test set. Uses output from AMARETTO_Run() and CreateRegulatorData().

Usage

```
AMARETTO_EvaluateTestSet(AMARETTOresults = AMARETTOresults,  
  MA_Data_TestSet = MA_Data_TestSet,  
  RegulatorData_TestSet = RegulatorData_TestSet)
```



```
AMARETTOresults <- AMARETTO_Run(AMARETTOinit)

AMARETTO_ExportResults(AMARETTOinit,AMARETTOresults,"./")
```

```
AMARETTO_HTMLreport  AMARETTO_HTMLreport
```

Description

Retrieve an interactive html report, including gene set enrichment analysis if asked for.

Usage

```
AMARETTO_HTMLreport(AMARETTOinit, AMARETTOresults, CNV_matrix = NULL,
  MET_matrix = NULL, SAMPLE_annotation = NULL, ID = NULL,
  VarPercentage, hyper_geo_test_bool = FALSE,
  hyper_geo_reference = NULL, output_address = "./", MSIGDB = FALSE,
  GMTURL = FALSE)
```

Arguments

AMARETTOinit	AMARETTO initialize output
AMARETTOresults	AMARETTO results output
CNV_matrix	Processed CNV matrix (\$CNV_matrix element from Preprocess_CancerSite() list output).
MET_matrix	Processed methylation matrix (\$MET_matrix element from Preprocess_CancerSite() list output).
SAMPLE_annotation	SAMPLE annotation will be added to heatmap
ID	ID column of the SAMPLE annotation data frame
VarPercentage	Original Var Percentage used
hyper_geo_test_bool	Boolean if a hyper geometric test needs to be performed. If TRUE provide a GMT file in the hyper_geo_reference parameter.
hyper_geo_reference	GMT file with gene sets to compare with.
output_address	Output directory for the html files.
MSIGDB	TRUE if gene sets were retrieved from MSIGDB. Links will be created in the report.
GMTURL	TRUE if second column of gmt contains URLs to gene set, FALSE if it contains a description

Examples

```
data("ProcessedDataLIHC")
AMARETTOinit <- AMARETTO_Initialize(MA_matrix = ProcessedDataLIHC$MA_matrix,
                                   CNV_matrix = ProcessedDataLIHC$CNV_matrix,
                                   MET_matrix = ProcessedDataLIHC$MET_matrix,
                                   NrModules = 10, VarPercentage = 5)

AMARETTOresults <- AMARETTO_Run(AMARETTOinit)

AMARETTO_HTMLreport(AMARETTOinit= AMARETTOinit,AMARETTOresults= AMARETTOresults,CNV_matrix=ProcessedDataLIHC$CNV_matrix,
                    MET_matrix = ProcessedDataLIHC$MET_matrix,VarPercentage=10,hyper_geo_test_bool=FALSE,output_ad
```

AMARETTO_Initialize *AMARETTO_Initialize (version: reorder and filter MA_Matrix)*

Description

Code used to initialize the seed clusters for an AMARETTO run. Requires processed gene expression (rna-seq or microarray), CNV (usually from a GISTIC run), and methylation (from MethylMix, provided in this package) data. Uses the function CreateRegulatorData() and results are fed into the function AMARETTO_Run().

Usage

```
AMARETTO_Initialize(MA_matrix = MA_matrix, CNV_matrix = NULL,
                    MET_matrix = NULL, Driver_list = NULL, NrModules, VarPercentage,
                    PvalueThreshold = 0.001, RsquareThreshold = 0.1, pmax = 10,
                    NrCores = 1, OneRunStop = 0, method = "union")
```

Arguments

MA_matrix	Expression matrix, with genes in rows and samples in columns.
CNV_matrix	CNV matrix, with genes in rows and samples in columns.
MET_matrix	Methylation matrix, with genes in rows and samples in columns.
Driver_list	Custom list of driver genes to be considered in analysis
NrModules	How many gene co-expression modules should AMARETTO search for? Usually around 100 is acceptable, given the large number of possible driver-passenger gene combinations.
VarPercentage	Minimum percentage by variance for filtering of genes; for example, 75% would indicate that the CreateRegulatorData() function only analyses genes that have a variance above the 75th percentile across all samples.
PvalueThreshold	Threshold used to find relevant driver genes with CNV alterations: maximal p-value.
RsquareThreshold	Threshold used to find relevant driver genes with CNV alterations: minimal R-square value between CNV and gene expression data.

pmax	"pmax" variable for glmnet function from glmnet package; the maximum number of variables aver to be nonzero. Should not be changed by user unless she/he fully understands the AMARETTO algorithm and how its parameters choices affect model output.
NrCores	A numeric variable indicating the number of computer/server cores to use for parallelization. Default is 1, i.e. no parallelization. Please check your computer or server's computing capacities before increasing this number. Parallelization is done via the RParallel package. Mac vs. Windows environments may behave differently when using parallelization.
method	Perform union or intersection of the driver genes evaluated from the input data matrices and custom driver gene list provided.

Examples

```
data("ProcessedDataLIHC")
data("Driver_Genes")
AMARETTOinit <- AMARETTO_Initialize(MA_matrix = ProcessedDataLIHC$MA_matrix,
                                   CNV_matrix = ProcessedDataLIHC$CNV_matrix,
                                   MET_matrix = ProcessedDataLIHC$MET_matrix,
                                   NrModules = 10, VarPercentage = 5)

AMARETTOinit <- AMARETTO_Initialize(MA_matrix = ProcessedDataLIHC$MA_matrix,
                                   CNV_matrix = NULL,
                                   MET_matrix = ProcessedDataLIHC$MET_matrix,
                                   Driver_list = Driver_Genes[["MSigDB"]],
                                   NrModules = 10, VarPercentage = 5)
```

AMARETTO_Preprocess *AMARETTO_Preprocess*

Description

Wrapper code that analyzes process TCGA GISTIC (CNV) and gene expression (rna-seq or microarray) data via one call

Usage

```
AMARETTO_Preprocess(CancerSite, DataSetDirectories)
```

Arguments

DataSetDirectories

Examples

```
CancerSite <- "LIHC"
TargetDirectory <- "../Downloads/" # path to data download directory
DataSetDirectories <- AMARETTO_Download(CancerSite,TargetDirectory)
ProcessedData <- AMARETTO_Preprocess(CancerSite,DataSetDirectories)
```

AMARETTO_Run

AMARETTO_Run

Description

Function to run AMARETTO, a statistical algorithm to identify cancer drivers by integrating a variety of omics data from cancer and normal tissue.

Usage

```
AMARETTO_Run(AMARETTOinit)
```

Arguments

AMARETTOinit List output from AMARETTO_Initialize().

Examples

```
data("ProcessedDataLIHC")
AMARETTOinit <- AMARETTO_Initialize(MA_matrix = ProcessedDataLIHC$MA_matrix,
                                     CNV_matrix = ProcessedDataLIHC$CNV_matrix,
                                     MET_matrix = ProcessedDataLIHC$MET_matrix,
                                     NrModules = 10, VarPercentage = 5)

AMARETTOresults <- AMARETTO_Run(AMARETTOinit)
```

AMARETTO_VisualizeModule

AMARETTO_VisualizeModule

Description

Function to visualize the gene modules

Usage

```
AMARETTO_VisualizeModule(AMARETTOinit, AMARETTOresults,
                          CNV_matrix = NULL, MET_matrix = NULL, ModuleNr,
                          SAMPLE_annotation = NULL, ID = NULL, order_samples = NULL)
```

Arguments

AMARETTOinit List output from AMARETTO_Initialize().
AMARETTOresults List output from AMARETTO_Run().
CNV_matrix CNV matrix, with genes in rows and samples in columns.
MET_matrix Methylation matrix, with genes in rows and samples in columns.
ModuleNr Module number to visualize.
SAMPLE_annotation Matrix or Dataframe with sample annotation.
ID Column used as sample name.
order_samples Order samples in heatmap by mean or by clustering.

Arguments

MA_matrix	Processed gene expression data matrix (\$MA_matrix element from Preprocess_CancerSite() list output).
CNV_matrix	Processed CNV matrix (\$CNV_matrix element from Preprocess_CancerSite() list output).
MET_matrix	Processed methylation matrix (\$MET_matrix element from Preprocess_CancerSite() list output).
Driver_list	Custom list of driver genes.
PvalueThreshold	Threshold used to find relevant driver genes with CNV alterations: maximal p-value.
RsquareThreshold	Threshold used to find relevant driver genes with CNV alterations: minimal R-square value between CNV and gene expression data.
method	Perform union or intersection of the driver genes evaluated from the input data matrices and custom driver gene list provided.

Examples

```
CreateRegulatorData(MA_matrix, CNV_matrix, MET_matrix, PvalueThreshold = 0.001, RsquareThreshold = 0.1)
```

Driver_Genes	<i>Driver_Genes</i>
--------------	---------------------

Description

A list of cancer driver genes described in literature.

Usage

```
Driver_Genes
```

Format

An object of class list of length 10.

Source

AMARETTO

MethylStates	<i>MethylStates</i>
--------------	---------------------

Description

A list of dataframes of Methyl States for cancer types obtained from MethylMix processing of TCGA methylation data

Usage

```
MethylStates
```

Format

An object of class `list` of length 25.

Source

AMARETTO

ProcessedDataLIHC	<i>ProcessedDataLIHC</i>
-------------------	--------------------------

Description

A list of dataframes of processed toy example dataset from TCGA-LIHC.

Usage

```
ProcessedDataLIHC
```

Format

An object of class `list` of length 3.

Source

AMARETTO

Index

*Topic **datasets**

- BatchData, [8](#)
- Driver_Genes, [9](#)
- MethylStates, [10](#)
- ProcessedDataLIHC, [10](#)

- AMARETTO_Download, [2](#)
- AMARETTO_EvaluateTestSet, [2](#)
- AMARETTO_ExportResults, [3](#)
- AMARETTO_HTMLreport, [4](#)
- AMARETTO_Initialize, [5](#)
- AMARETTO_Preprocess, [6](#)
- AMARETTO_Run, [7](#)
- AMARETTO_VisualizeModule, [7](#)

- BatchData, [8](#)

- CreateRegulatorData, [8](#)

- Driver_Genes, [9](#)

- MethylStates, [10](#)

- ProcessedDataLIHC, [10](#)