Package 'AMARETTO'

February 11, 2019

Type Package

Title Regulatory network inference and driver gene evaluation using integrative multi-omics analysis and penalized regression.

Version 0.99.4

Date 2016-06-06

Author Jayendra Shinde, Celine Everaert, Shaimaa Bakr, Mohsen Nabian
Nathalie Pochet and Olivier Gevaert

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

Depends R (>= 3.5), RCurl,limma, foreach, parallel, doParallel,
glmnet (>= 2.0.16), 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,Preproce

Suggests testthat,
MASS,
knitr,
rmarkdown

NeedsCompilation no

Imports callr (>= 3.0.0.9001),
Rcpp

Roxygen list(markdown = TRUE)

RoxygenNote 6.1.1

LinkingTo Rcpp

VignetteBuilder knitr

R topics documented:

AMARETTO_Download   *AMARETTO_Download*

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

AMARETTO_EvaluateTestSet

AMARETTO_EvaluateTestSet

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)
```

Arguments

AMARETTOresults

                AMARETTO output from AMARETTO_Run().

MA_Data_TestSet

                Gene expression matrix from a test set (that was not used in AMARETTO_Run()).

RegulatorData_TestSet

                Test regulator data from CreateRegulatorData().

Examples

```
data('ProcessedDataLIHC')
AMARETTOinit <- AMARETTO_Initialize(MA_matrix = ProcessedDataLIHC$MA_matrix,
                                    CNV_matrix = ProcessedDataLIHC$CNV_matrix,
                                    MET_matrix = ProcessedDataLIHC$MET_matrix,
                                    NrModules = 20, VarPercentage = 50)

AMARETTOresults <- AMARETTO_Run(AMARETTOinit)

AMARETTOtestReport <- AMARETTO_EvaluateTestSet(AMARETTOresults = AMARETTOresults,
                                    MA_Data_TestSet = AMARETTOinit$MA_matrix_Var,
                                  RegulatorData_TestSet = AMARETTOinit$RegulatorData)
```

AMARETTO_ExportResults

AMARETTO_ExportResults

Description

Retrieve a download of all the data linked with the run (including heatmaps)

Usage

```
AMARETTO_ExportResults(AMARETTOinit, AMARETTOresults, data_address,
  Heatmaps = TRUE, CNV_matrix = NULL, MET_matrix = NULL)
```

Arguments

`AMARETTOinit`	AMARETTO initialize output
`AMARETTOresults`	
	AMARETTO results output
`data_address`	Directory to save data folder
`Heatmaps`	Output heatmaps as pdf

Examples

```
data('ProcessedDataLIHC')
AMARETTOinit <- AMARETTO_Initialize(MA_matrix = ProcessedDataLIHC$MA_matrix,
                                    CNV_matrix = ProcessedDataLIHC$CNV_matrix,
                                    MET_matrix = ProcessedDataLIHC$MET_matrix,
                                    NrModules = 20, VarPercentage = 50)

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 = 20, VarPercentage = 50)

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_address='./')
```

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 expres-
siosn (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? Usu-
              ally 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 num-
                 ber 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
                 paralellelization. Default is 1, i.e. no parallelization. Please check your com-
                 puter or server's computing capacities before increasing this number. Paral-
                 lelization 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 = 20, VarPercentage = 50)

AMARETTOinit <- AMARETTO_Initialize(MA_matrix = ProcessedDataLIHC$MA_matrix,
                                    CNV_matrix = NULL,
                                    MET_matrix = ProcessedDataLIHC$MET_matrix,
                                    Driver_list = Driver_Genes[['MSigDB']],
                                    NrModules = 20, VarPercentage = 50)
```

AMARETTO_Preprocess   *AMARETTO_Preprocess*

Description

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

Usage

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

Arguments

DataSetDirectories

Examples

```
data('BatchData')
TargetDirectory <- file.path(getwd(),'Downloads/') # path to data download directory
CancerSite <- 'LIHC'
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 = 20, VarPercentage = 50)

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.

Examples

```
data('ProcessedDataLIHC')
AMARETTOinit <- AMARETTO_Initialize(MA_matrix = ProcessedDataLIHC$MA_matrix,
                                    CNV_matrix = ProcessedDataLIHC$CNV_matrix,
                                    MET_matrix = ProcessedDataLIHC$MET_matrix,
                                    NrModules = 20, VarPercentage = 50)

AMARETTOresults <- AMARETTO_Run(AMARETTOinit)

ModuleNr <- 1 #define the module number to visualize

AMARETTO_VisualizeModule(AMARETTOinit = AMARETTOinit,AMARETTOresults = AMARETTOresults,
                         CNV_matrix = ProcessedDataLIHC$CNV,
                         MET_matrix = ProcessedDataLIHC$MET,
                         ModuleNr = ModuleNr)
```

BatchData	*BatchData*

Description

A dataset for conducting batch corerction in TCGA samples

Usage

```
BatchData
```

Format

An object of class `data.frame` with 23263 rows and 3 columns.

Source

[AMARETTO](#)

Driver_Genes *Driver_Genes*

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 *MethylStates*

Description

A list of dataframes of Methyl States for example dataset from TCGA-LIHC.

Usage

MethylStates

Format

An object of class list of length 1.

Source

AMARETTO

ProcessedDataLIHC   *ProcessedDataLIHC*

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