Package: AMARETTO (via r-universe)

June 14, 2026

Type Package

Title Regulatory Network Inference and Driver Gene Evaluation using Integrative Multi-Omics Analysis and Penalized Regression

Version 1.28.0

Date 2016-06-06

Author Jayendra Shinde, Celine Everaert, Shaimaa Bakr, Mohsen Nabian, Jishu Xu, Vincent Carey, Nathalie Pochet and Olivier Gevaert

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

Depends R (>= 3.6), impute, doParallel, grDevices, dplyr, methods, ComplexHeatmap

Description Integrating an increasing number of available multi-omics cancer data remains one of the main challenges to improve our understanding of cancer. One of the main challenges is using multi-omics data for identifying novel cancer driver genes. We have developed an algorithm, called AMARETTO, that integrates copy number, DNA methylation and gene expression data to identify a set of driver genes by analyzing cancer samples and connects them to clusters of co-expressed genes, which we define as modules. We applied AMARETTO in a pancancer setting to identify cancer driver genes and their modules on multiple cancer sites. AMARETTO captures modules enriched in angiogenesis, cell cycle and EMT, and modules that accurately predict survival and molecular subtypes. This allows AMARETTO to identify novel cancer driver genes directing canonical cancer pathways.

License Apache License (== 2.0) + file LICENSE

LazyLoad yes

LazyData true

Encoding UTF-8

biocViews
StatisticalMethod,DifferentialMethylation,GeneRegulation,GeneExpression,MethylationArray,Transcription,Preprocessin

Suggests testthat, MASS, knitr, BiocStyle

NeedsCompilation no

Imports callr (>= 3.0.0.9001), Matrix, Rcpp, BiocFileCache, DT, MultiAssayExperiment, circlize, curatedTCGAData, foreach, glmnet, httr, limma, matrixStats, readr, reshape2, tibble, rmarkdown, graphics, grid, parallel, stats, knitr, ggplot2, gridExtra, utils

Roxygen list(markdown = TRUE)

RoxygenNote 6.1.1.9000

LinkingTo Rcpp

VignetteBuilder knitr

Config/pak/sysreqs
cmake make libicu-dev libpng-dev libuv1-dev libssl-dev perl libx11-dev zlib1g-dev

Repository https://bioc-release.r-universe.dev

Date/Publication 2026-04-28 12:49:47 UTC

RemoteUrl https://github.com/bioc/AMARETTO

RemoteRef RELEASE_3_23

RemoteSha 4b60a9dafc2681e6700f8ba6aef3a05cd42275e1

Contents

`AMARETTO_CreateModuleData` *AMARETTO_CreateModuleData*

Description

AMARETTO_CreateModuleData

Usage

```
AMARETTO_CreateModuleData(AMARETTOinit, AMARETTOresults)
```

Arguments

`AMARETTOinit`	List output from AMARETTO_Initialize().
`AMARETTOresults`	List output from AMARETTO_Run()

Value

result

Examples

```
data('ProcessedDataLIHC')
AMARETTOinit <- AMARETTO_Initialize(ProcessedData = ProcessedDataLIHC,
                                    NrModules = 2, VarPercentage = 50)
AMARETTOresults <- AMARETTO_Run(AMARETTOinit)
AMARETTO_MD <- AMARETTO_CreateModuleData(AMARETTOinit, AMARETTOresults)
```

`AMARETTO_CreateRegulatorPrograms` *AMARETTO_CreateRegulatorPrograms*

Description

AMARETTO_CreateRegulatorPrograms

Usage

```
AMARETTO_CreateRegulatorPrograms(AMARETTOinit, AMARETTOresults)
```

Arguments

`AMARETTOinit`	List output from AMARETTO_Initialize().
`AMARETTOresults`	List output from AMARETTO_Run()

Value

result

Examples

```
data('ProcessedDataLIHC')
AMARETTOinit <- AMARETTO_Initialize(ProcessedData = ProcessedDataLIHC,
                                    NrModules = 2, VarPercentage = 50)
AMARETTOresults <- AMARETTO_Run(AMARETTOinit)
AMARETTO_RP <- AMARETTO_CreateRegulatorPrograms(AMARETTOinit,AMARETTOresults)
```

`AMARETTO_Download` *AMARETTO_Download*

Description

Downloading TCGA dataset for AMARETTO analysis

Usage

```
AMARETTO_Download(CancerSite = "CHOL",
  TargetDirectory = TargetDirectory)
```

Arguments

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

Value

result

Examples

```
TargetDirectory <- file.path(getwd(),"Downloads/");dir.create(TargetDirectory)
CancerSite <- 'CHOL'
DataSetDirectories <- AMARETTO_Download(CancerSite,TargetDirectory = 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().

Value

result

Examples

```
data('ProcessedDataLIHC')
AMARETTOinit <- AMARETTO_Initialize(ProcessedData = ProcessedDataLIHC,
                                    NrModules = 2, 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
`CNV_matrix`	CNV_matrix
`MET_matrix`	MET_matrix

Value

result

Examples

```
data('ProcessedDataLIHC')
TargetDirectory <- file.path(getwd(),"Downloads/");dir.create(TargetDirectory)
AMARETTOinit <- AMARETTO_Initialize(ProcessedData = ProcessedDataLIHC,
                                    NrModules = 2, VarPercentage = 50)

AMARETTOresults <- AMARETTO_Run(AMARETTOinit)
AMARETTO_ExportResults(AMARETTOinit,AMARETTOresults,TargetDirectory,Heatmaps = FALSE)
```

`AMARETTO_HTMLreport` *AMARETTO_HTMLreport*

Description

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

Usage

```
AMARETTO_HTMLreport(AMARETTOinit, AMARETTOresults, ProcessedData,
  show_row_names = FALSE, SAMPLE_annotation = NULL, ID = NULL,
  hyper_geo_test_bool = FALSE, hyper_geo_reference = NULL,
  output_address = "./", MSIGDB = TRUE, driverGSEA = TRUE,
  phenotype_association_table = NULL)
```

Arguments

`AMARETTOinit`	AMARETTO initialize output
`AMARETTOresults`	AMARETTO results output
`ProcessedData`	List of processed input data
`show_row_names`	if True, sample names will appear in the heatmap
`SAMPLE_annotation`	SAMPLE annotation will be added to heatmap
`ID`	ID column of the SAMPLE annotation data frame
`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.
`driverGSEA`	if TRUE, module drivers will also be included in the hypergeometric test.
`phenotype_association_table`	a Data Frame, containing all modules phenotype association data. Optional.

Value

result

Examples

```
## Not run:
data('ProcessedDataLIHC')
AMARETTOinit <- AMARETTO_Initialize(ProcessedData = ProcessedDataLIHC,
                                    NrModules = 2, VarPercentage = 50)

AMARETTOresults <- AMARETTO_Run(AMARETTOinit)

AMARETTO_HTMLreport(AMARETTOinit= AMARETTOinit,AMARETTOresults= AMARETTOresults,
                    ProcessedData = ProcessedDataLIHC,
                    hyper_geo_test_bool=FALSE,
                    output_address='./')

## End(Not run)
```

`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 expressiosn (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(ProcessedData = ProcessedData, Driver_list = NULL,
  NrModules, VarPercentage, PvalueThreshold = 0.001,
  RsquareThreshold = 0.1, pmax = 10, NrCores = 1, OneRunStop = 0,
  method = "union", random_seeds = NULL, convergence_cutoff = 0.01)
```

Arguments

`ProcessedData`	List of Expression, CNV and MethylMix data matrices, 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 paralellelization. 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.
`OneRunStop`	OneRunStop
`method`	Perform union or intersection of the driver genes evaluated from the input data matrices and custom driver gene list provided.
`random_seeds`	A numeric vector of length 2, containing two seed numbers for randomization : 1st for kmeans and 2nd for glmnet
`convergence_cutoff`	A numeric value (E.g. 0.01) representing the fraction of the total number of genes, in which, The algorithm is considered reaching convergence and will stop, if Nr of Gene-replacements in an iteration falls below this threshold * total number of genes.

Value

result

Examples

```
data('ProcessedDataLIHC')
data('Driver_Genes')
AMARETTOinit <- AMARETTO_Initialize(ProcessedData = ProcessedDataLIHC,
                                    NrModules = 2, VarPercentage = 50)
## Not run:
AMARETTOinit <- AMARETTO_Initialize(ProcessedData = ProcessedDataLIHC,
                                    Driver_list = Driver_Genes[['MSigDB']],
                                    NrModules = 2, VarPercentage = 50)

## End(Not run)
```

`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(DataSetDirectories = DataSetDirectories,
  BatchData = BatchData)
```

Arguments

`DataSetDirectories`	DataSetDirectories
`BatchData`	BatchData

Value

result

Examples

```
## Not run:
TargetDirectory <-  "Downloads"  # path to data download directory
CancerSite <- 'CHOL'
DataSetDirectories <- AMARETTO_Download(CancerSite,TargetDirectory)
ProcessedData <- AMARETTO_Preprocess(DataSetDirectories,BatchData)

## End(Not run)
```

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

Description

AMARETTO_Run 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().

Value

result

Examples

```
data('ProcessedDataLIHC')
AMARETTOinit <- AMARETTO_Initialize(ProcessedData = ProcessedDataLIHC,
                                    NrModules = 2, VarPercentage = 50)
AMARETTOresults <- AMARETTO_Run(AMARETTOinit)
```

`AMARETTO_VisualizeModule`

AMARETTO_VisualizeModule

Description

Function to visualize the gene modules

Usage

```
AMARETTO_VisualizeModule(AMARETTOinit, AMARETTOresults, ProcessedData,
  ModuleNr, show_row_names = FALSE, SAMPLE_annotation = NULL,
  ID = NULL, order_samples = NULL)
```

Arguments

`AMARETTOinit`	List output from AMARETTO_Initialize().
`AMARETTOresults`	List output from AMARETTO_Run().
`ProcessedData`	List of processed input data
`ModuleNr`	Module number to visualize
`show_row_names`	If TRUE, row names will be shown on the plot.
`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

Value

result

Examples

```
data('ProcessedDataLIHC')
AMARETTOinit <- AMARETTO_Initialize(ProcessedData = ProcessedDataLIHC,
                                    NrModules = 2, VarPercentage = 50)

AMARETTOresults <- AMARETTO_Run(AMARETTOinit)

AMARETTO_VisualizeModule(AMARETTOinit = AMARETTOinit,AMARETTOresults = AMARETTOresults,
                         ProcessedData = ProcessedDataLIHC, ModuleNr = 1)
```

`BatchData` *BatchData*

Description

A dataset for conducting batch corerction in TCGA samples

Usage

```
BatchData
```

Format

A data frame with 23263 observations and 3 variables:

Source

`AMARETTO`

`Driver_Genes` *Driver_Genes*

Description

A list of cancer driver genes described in literature.

Usage

```
Driver_Genes
```

Format

List

Source

`AMARETTO`

`MsigdbMapping` *MsigdbMapping*

Description

A dataset containing all MSIGDB pathways and their descriptions. .

Usage

```
MsigdbMapping
```

Format

List

Source

`AMARETTO`

`plot_run_history` *Title plot_run_history*

Description

Title plot_run_history

Usage

```
plot_run_history(AMARETTOinit, AMARETTOresults)
```

Arguments

`AMARETTOinit`	AMARETTO initialize output
`AMARETTOresults`	AMARETTO results output

Value

plot

Examples

```
data('ProcessedDataLIHC')
AMARETTOinit <- AMARETTO_Initialize(ProcessedData = ProcessedDataLIHC,
                                    NrModules = 2, VarPercentage = 50)

AMARETTOresults <- AMARETTO_Run(AMARETTOinit)

plot_run_history(AMARETTOinit,AMARETTOresults)
```

`ProcessedDataLIHC` *ProcessedDataLIHC*

Description

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

Usage

```
ProcessedDataLIHC
```

Format

List

Source

`AMARETTO`

`read_gct` *read_gct*

Description

Function to turn a .gct data files into a matrix format

Usage

```
read_gct(file_address)
```

Arguments

`file_address`	Address of the input gct file.

Value

result

Examples

```
data_matrix<-read_gct(file_address="")
```

Index