Package ‘pbcmc’

November 12, 2016

Type Package

Title Permutation-Based Confidence for Molecular Classification

Version 1.2.0

Date 2016-09-23

Author Cristobal Fresno, German A. Gonzalez, Andrea S. Llera and Elmer A. Fernandez

Maintainer Cristobal Fresno <cristobalfresno@gmail.com>

Description The pbcmc package characterizes uncertainty assessment on gene expression classifiers, a. k. a. molecular signatures, based on a permutation test. In order to achieve this goal, synthetic simulated subjects are obtained by permutations of gene labels. Then, each synthetic subject is tested against the corresponding subtype classifier to build the null distribution. Thus, classification confidence measurement can be provided for each subject, to assist physician therapy choice. At present, it is only available for PAM50 implementation in genefu package but it can easily be extend to other molecular signatures.

URL http://www.bdmg.com.ar/

License GPL (>=2)

Depends R (>= 3.3.0), genefu

Imports Biobase, BiocGenerics, BiocParallel (>= 1.3.13), parallel, reshape2, grid, utils, cowplot, methods, limma, ggplot2, gridExtra, grDevices, stats

Suggests breastCancerUPP, breastCancerNKI, breastCancerVDX, breastCancerTRANSBIG, breastCancerMAINZ, breastCancerUNT

biocViews Classification, GeneExpression, Microarray, MultipleComparison, QualityControl, Normalization, Clustering, mRNAMicroarray, OneChannel, TwoChannel, RNASeq, KEGG, DifferentialExpression

as PAM50 high level coerce functions

Description
These functions (setAs and as.PAM50) are intended to be used with limma MAList-class in order to coerce its structure into a compatible PAM50 class.

Usage
as(object,Class,strict=TRUE,ext=possibleExtends(thisClass,Class))
as.PAM50(object)

## S4 method for signature 'MAList'
as.PAM50(object)

Arguments
- object: MAList object with at least M and genes items, optionally targets.
- Class: character with the name of class "PAM50" to be coerced.
- strict, ext: see as function.
Details

Basically the $M and $genes items are copied into a MolecularPermutationClassifier’s exprs and annotation slots respectively. In addition, if present, $targets content is also copied to the same named slot.

Value

a PAM50 object with the respective copied data.

Author(s)

Cristobal Fresno <cfresno@bdmg.com.ar>, German A. Gonzalez <ggonzalez@bdmg.com.ar>, Andrea S. Llera <alleraleloir.org.ar> and Elmer Andres Fernandez <efernandez@bdmg.com.ar>

See Also

PAM50 for a complete example.


Examples

```r
##Example 1: Create a PAM50 object -----------------------------------------
##1) Just an empty object
object<-PAM50()
object

##2) Using Breast Cancer NKI database, if available.
if(requireNamespace("breastCancerNKI")){
  object<-loadBCDataset(Class=PAM50, libname="nki", verbose=TRUE)
  object
  ##Now we can inspect the object
  head(exprs(object))  ##The gene expression
  head(annotation(object))  ##The available annotation
  head(targets(object))  ##The clinical data present in the package
}

##Example 2: Build a PAM50 object with user data --------------------------
##Option 1: using PAM50 constructor. The user will only need:
##a) The M gene expression object, i. e., gene in rows and sample in columns
##b) The annotation data.frame which must include the compulsory fields
## "probe", "NCBI.gene.symbol" and "EntrezGene.ID"
M<-pam50$centroids
genes<-pam50$centroids.map
names(genes)<-c("probe", "NCBI.gene.symbol", "EntrezGene.ID")
object<-PAM50(exprs=M, annotation=genes)
object

##Option 2: Two ways to build it from a MAList (as or as.PAM50)-------------
##Let's use PAM50 classifier's centroids toy example, i. e., the five subject
##subtypes, which must correctly classify all the subject.
M<-pam50$centroids
genes<-pam50$centroids.map
names(genes)<-c("probe", "NCBI.gene.symbol", "EntrezGene.ID")
maux<-new("MAList", list(M=M, genes=genes))
```
classify,PAM50-method
classify subjects with PAM50 molecular signature

Description
Obtain PAM50 subtype using genefu centroid Spearman’s correlation implementation. If std="median" probes with the same mapping are averaged. Then, the complete database is center normalized using gene median expression. This is done in order to assure selecting the same "gene" to those in "genefu" library, instead of the most variant probe (default in geneid.map), when more than one probe match the same gene. This selection is based on probe population variance that could depends on the number of accounted genes.

Usage
## S4 method for signature 'PAM50'
classify(object, std = c("none", "scale", "robust", "median") [1], verbose = getOption("verbose", default = FALSE))

Arguments
- **object**: a MolecularPermutationClassifier subclass object.
- **std**: character to select standardization alternative "none" (default), "scale" and "robust" as in genefu original implementation, plus the suggested "median" if many subjects are available.
- **verbose**: should the user feedback be displayed? By default value is "verbose" global option parameter, if present, or FALSE otherwise.

Value
a PAM50 object with the updated slots:
- **@exprs**: updated matrix with the used std parameter.
- **@classification**
  - **$subtype**: subject named factor with all classifier possible levels, i.e, "Basal", "Her2", "LumA", "LumB" and "Normal".
  - **$probability**: numeric matrix with subtype class probability for each subject, as in genefu, obtained as the positive proportion of correlation explained by each subtype.
  - **$correlation**: numeric matrix with Spearman’s rho correlation of each subject to the corresponding PAM50 subtypes.
Author(s)

Cristobal Fresno <cfresno@bdmg.com.ar>, German A. Gonzalez <ggonzalez@bdmg.com.ar>, Andrea S. Llera <allera@leloir.org.ar> and Elmer Andres Fernandez <efernandez@bdmg.com.ar>

References


See Also

PAM50 for a complete example.

Other PAM50: as, filtrate,PAM50-method, pam50centroids, permutate, PAM50-method, subjectReport, PAM50-method

Examples

```r
## Using pam50centroids package example data
data(pam50centroids)

## Get the original PAM50 calls using genefu implementation
pam50centroids<-classify(pam50centroids, std="none", verbose=TRUE)
classification(pam50centroids)
```

filtrate

Virtual functions for MolecularPermutationClassifier hierarchy

Description

The following functions establish an organized framework for MolecularPermutationClassifier sub-classes data processing. In this context, the later are supposed to be implemented with respective responsibilities. In particular, once the class is created the user has to:

filtrate: Removes, from the exprs matrix, subjects not required by the classification algorithm.
classify: Generates subject classification according to subclass implementations (PAM50, etc.).
permutate: Obtains subject classification based on the null correlation distribution by means permutation simulation.
subtype: Obtain the new classification using permutation results.
databaseReport: A pdf with all subjectReports, if a database is available.
Usage

filtrate(object, verbose = getOption("verbose", default = FALSE))

classify(object, ..., verbose = getOption("verbose", default = FALSE))

permutate(object, nPerm = 10000L, pCutoff = 0.01, where = "fdr",

    keep = FALSE, ..., seed = 1234567890, BPPARAM = bpparam(),

    verbose = getOption("verbose", default = TRUE))

subtypes(object, pCutoff = 0.01, ..., where = c("fdr", "pvalue")[1])

subjectReport(object, subject)

databaseReport(object, fileName, ..., verbose = getOption("verbose", default = TRUE))

Arguments

object MolecularPermutationClassifier child class object

verbose should the user feedback be displayed? By default value is "verbose" global

option parameter, if present, or FALSE otherwise.

... additional parameters for future implementations.

nPerm integer with number of permutations. Default: 1e4L.

pCutoff numeric with p-value or fdr cutoff used, i.e., variable<pCutoff. Default: 0.01.

where character with significant value used. Default value is "fdr".

keep should null distribution simulation values be kept?. Default: FALSE

seed integer to use as random seed. Default: 1234567890.

BPPARAM an optional BiocParallelParam instance determining the parallel back-end to be

used during evaluation, or a list of BiocParallelParam instances, to be applied in

sequence for nested calls to bplapply. Default=bpparam().

subject integer to select the appropriate subject to report.

fileName character with the name of the pdf report file to save.

Value

A MolecularPermutationClassifier child according to the actual object class.

Author(s)

Cristobal Fresno <cfresno@bdmg.com.ar>, German A. Gonzalez <ggonzalez@bdmg.com.ar>,

Andrea S. Llera <allera@leloir.org.ar> and Elmer Andres Fernandez <efernandez@bdmg.com.ar>

See Also

PAM50 for a complete example.

Other MolecularPermutationClassifier PAM50: PAM50-class, loadBCDataset
Examples

```r
## Using pam50centroids package example data
data(pam50centroids)
pam50centroids
pam50centroids <- filtrate(pam50centroids, verbose=TRUE)
pam50centroids <- classify(pam50centroids, std="none", verbose=TRUE)
## Let's run a quick example with 100 permutations. It is recommended at
## least 10,000
pam50centroids <- permutate(pam50centroids, nPerm=100, pCutoff=0.01,
corCutoff=0.1, verbose=TRUE)
pam50centroids
```

filtrate,PAM50-method
filtrate centroid genes from PAM50 classification

Description

Remove exprs rows not required by MolecularPermutationClassifier subclasses to classify samples, in this case PAM50. This means to only keep genes with valid EntrezGeneID, i.e., not NA and present in PAM50 signature centroids. In addition, annotation slot will only keep "probe", "EntrezGene.ID" and "NCBI.gene.symbol" fields required by genefu’s intrinsic.cluster.predict function.

Usage

```r
## S4 method for signature 'PAM50'
filtrate(object, verbose = getOption("verbose", default = FALSE))
```

Arguments

- `object` a PAM50 object.
- `verbose` should the user feedback be displayed? By default value is "verbose" global option parameter, if present, or FALSE otherwise.

Value

MolecularPermutationClassifier subclass with updated slots:
- `@exprs` only rows required by the classifier.
- `@annotation` consistent with exprs rows and only "probe", "EntrezGene.ID" and "NCBI.gene.symbol" annotation fields.

Author(s)

Cristobal Fresno <cfresno@bdmg.com.ar>, German A. Gonzalez <ggonzalez@bdmg.com.ar>, Andrea S. Llera <allera@leloir.org.ar> and Elmer Andres Fernandez <efernandez@bdmg.com.ar>

See Also

- `PAM50` for a complete example.
- Other PAM50: `as`, `classify,PAM50-method`, `pam50centroids`, `permutate,PAM50-method`, `subjectReport,PAM50-method`, `subtypes,PAM50-method`
Examples

```r
##Using pam50centroids package example data
data(pam50centroids)
pam50centroids
pam50centroids<-filtrate(pam50centroids, verbose=TRUE)
pam50centroids
##Using Breast Cancer NKI database, if available.
if(requireNamespace("breastCancerNKI")){
  object<-loadBCDataset(Class=PAM50, libname="nki", verbose=TRUE)
  object
  object<-filtrate(object, verbose=TRUE)
  object
}
```

loadBCDataset  

*MolecularPermutationClassifier high level constructor*

Description

High level constructor for MolecularPermutationClassifier subclasses using available Bioconductor's Breast Cancer example datasets.

Usage

```r
loadBCDataset(Class, libname = c("upp", "nki", "vdx", "mainz", "transbig", "unt"), verbose = getOption("verbose", default = FALSE))
```

```r
## S4 method for signature 'classGeneratorFunction'
loadBCDataset(Class, libname = c("upp", "nki", "vdx", "mainz", "transbig", "unt"), verbose = getOption("verbose", default = FALSE))
```

Arguments

- **Class**: name of MolecularPermutationClassifier child class to use.
- **libname**: lowercase character with the name of the breastCancerXXX database to be loaded. At present, XXX can be "upp", "nki", "vdx", "mainz", "transbig" or "unt". See reference for available breast cancer citations.
- **verbose**: should the user feedback be displayed? By default value is "verbose" global option parameter, if present, or FALSE otherwise.

Value

MolecularPermutationClassifier subclass object with exprs, annotation and targets slots taken from the libname used.

Author(s)

Cristobal Fresno <cfresno@bdmg.com.ar>, German A. Gonzalez <ggonzalez@bdmg.com.ar>, Andrea S. Llera <allera@leloir.org.ar> and Elmer Andres Fernandez <efernandez@bdmg.com.ar>
References


See Also

PAM50 for a complete example.

Other MolecularPermutationClassifier PAM50: PAM50-class, filtrate

Examples

```r
##Using Breast Cancer NKI database, if available, to create a PAM50 class.
if(requireNamespace("breastCancerNKI")){
  object<-loadBCDataset(Class=PAM50, libname="nki", verbose=TRUE)
  object

  #Now we can inspect the object
  head(exprs(object))  ##The gene expression
  head(annotation(object))  ##The available annotation
  head(targets(object))  ##The clinical data present in the package
}
```

MolecularPermutationClassifier-class

Class MolecularPermutationClassifier S4 implementation in R

Description

Virtual class to represent gene-based molecular signature classification by means of permutation test.
MolecularPermutationClassifier-class

Slots

parameters named list with at least the following fields:

  $nPerm  integer with number of permutations. Default: 1e4L
  $where  character with significant value used. Default value is "fdr".
  $pCutoff numeric with p-value or fdr cutoff used, i.e., variable<\text{pCutoff}. Default: 0.01
  $keep  should null distribution simulation values be kept?. Default: FALSE

exprs matrix with gene exprs profile, where genes are in rows and subjects as columns, a.k.a., M matrix.

annotation data.frame with individual annotations (genes, etc). Minimal compulsory fields are:

  $probe  same characters as in row.names(M).
  $EntrezGene.ID integer with NCBI Entrez Data Base.
  $NCBIgene.symbol character with gene mnemonic, a.k.a. gene symbol.

targets data.frame with additional subject data (optional).

classification named list with at least the following fields:

  $class  factor with all possible class levels.

permutation named list with at least the following fields:

  $pvalues numeric matrix with subjects in row and classes in columns.
  $fdr numeric matrix with False Discovery Rate correction of p-values by row.

Superclasses

None declared.

Subclasses


Functions

MolecularPermutationClassifier S4 class includes the following functions:

- Integrity check:
  
  validity will check appropriate annotation data.frame minimal required columns, all named parameters and if exprs and annotation dimension matches.

  prototype just for an empty class with default values: nPerm=1e4L, where="fdr", pCutoff=0.01, corCutoff=0.1 and keep=FALSE.

- Generics:
  
  show,print basic class display wrappers.

  summary classifier statistics.

- Constructors (as this class is virtual see subclass’ documentation).

  setAs MAList to PAM50

  as.PAM50 wrapper for PAM50 setAs from MAList.

  loadBCDataset wrapper to load BreastCancerXX data (Class, exprs, annotation, clinical data).

- Getters for the corresponding slots (parameters, exprs, annotation, targets, classification and permutation).
Setters for the corresponding slots (parameters<-, annotation<- and targets<-).

Particular (virtual) functions:

- **filtrate** remove from the exprs matrix subjects not required by the classification algorithm.
- **classify** generate subject classification according to subclasses implementation (PAM50, etc.).
- **permutate** obtain subject classification based on the null correlation distribution by means permutation simulation.
- **subtypes** obtain the new classification using permutation results.
- **subjectReport** a friendly report for Physician treatment decision support.
- **databaseReport** a pdf with all subjectReports, if a database is available.

Author(s)

Cristobal Fresno <cfresno@bdmg.com.ar>, German A. Gonzalez <ggonzalez@bdmg.com.ar>, Andrea S. Llera <allera@leloir.org.ar> and Elmer Andres Fernandez <efernandez@bdmg.com.ar>

References


See Also

- [PAM50](#) for a complete example, [loadBCDataset](#) to load BreastCancerXX dataset, **filtrate**, **classify** and **permutate** to get corresponding Breast Cancer subtype. Getters/Setters for this class are parameters, exprs, annotation, targets, classification and permutation.

Other MolecularPermutationClassifier: parameters, show

---

### PAM50-class

**PAM50 S4 implementation in R**

**Description**

This is a concrete MolecularPermutationClassifier based on Perou et al. (2000 & 2010) PAM50 molecular signature, using genefu package implementation (Haibe-Kains et al. 2014).

**Slots**

- **parameters** named list with at least the following fields:
  - $nPerm integer with number of permutations. Default: 1e4L
  - $where character with significant value used. Default value is "fdr".
  - $pCutoff numeric with p-value or fdr cutoff used, i.e., variable<pCutoff. Default: 0.01
  - $keep should null distribution simulation values be kept?. Default: FALSE
corCutoff  PAM50 additional numeric parameter with the correlation difference between classes
cutoff used, i.e., \( |\rho(profile, class_A) - \rho(profile, class_B)| > \text{corCutoff} \)

exprs  matrix with gene exprs profile, where genes are in rows and subjects as columns, a.k.a., M
matrix.

annotation  data.frame with individual annotations (genes, etc). Minimal compulsory fields are:
$probe  same characters as in row.names(M).
$EntrezGene.ID  integer with NCBI Entrez Data Base.
$NCBIgene.symbol  character with gene mnemonic, a.k.a. gene symbol.

targets  data.frame with additional subject data (optional).

classification  named list with at least the following fields:
$subtype  factor with PAM50 subtype of each sample.
$probability  matrix with the subtype probability of each subtype per sample, as in genefu
library.
$correlation  matrix with the observed correlation of each subtype per sample.

permutation  named list with at least the following fields:
$correlation  Only if keep==TRUE is a list of the five subtypes containing a matrix with the
permuted null distribution correlations.
$pvalues  matrix with the subject’s p-values of the permutation test per subject.
$fdr  matrix with the corresponding adjusted p-values.
$subtype  data.frame where each subject has the reported "PAM50" subtype, the "Permuted"
test result i.e. "Assigned", "Not Assigned" or "Ambiguous"; "Classes" whether is a single
PAM50 subtype or more than one if Ambiguous case; "Class" if it is needed to assign just
one i.e., a single PAM50 subtype or Not Assigned.

Superclasses

Direct descendant from MolecularPermutationClassifier-class.

Subclasses

None declared.

Function

Redefinition from MolecularPermutationClassifier: filtrate, classify, permutate, subjectReporta and
databaseReport.

Author(s)

Cristobal Fresno <cfresno@bdmg.com.ar>, German A. Gonzalez <ggonzalez@bdmg.com.ar>,
Andrea S. Llera <allera@leloir.org.ar> and Elmer Andres Fernandez <efernandez@bdmg.com.ar>

References

Relevant Functions for Gene Expression Analysis, Especially in Breast Cancer. R package
version 1.16.0, www.pmgenomics.ca/bhklab/
PAM50-class

See Also

Other MolecularPermutationClassifier PAM50: filtrate, loadBCDataset

Examples

##Example 1: Create a PAM50 object -----------------------------------------

###1) Just an empty object

```r
object<-PAM50()
```

###2) Using Breast Cancer NKI database, if available.

```r
if(requireNamespace("breastCancerNKI")){
  object<-loadBCDataset(Class=PAM50, libname="nki", verbose=TRUE)
  object
  #Now we can inspect the object
  head(exprs(object)) #The gene expression
  head(annotation(object)) #The available annotation
  head(targets(object)) #The clinical data present in the package
}
```

##Example 2: Build a PAM50 object with user data -------------------------

###1) using PAM50 constructor. The user will only need:

- The M gene expression object, i.e., gene in rows and sample in columns
- The annotation data.frame which must include the compulsory fields “probe”, “NCBI.gene.symbol” and “EntrezGene.ID”

```r
M<-pam50$centroids
genes<-pam50$centroids.map
names(genes)<-c("probe", "NCBI.gene.symbol", "EntrezGene.ID")
object<-PAM50(exprs=M, annotation=genes)
```

###2) Two ways to build it from a MAList (as or as.PAM50) -------------

Let’s use PAM50 classifier’s centroids toy example, i.e., the five subject subtypes, which must correctly classify all the subject.

```r
M<-pam50$centroids
genes<-pam50$centroids.map
names(genes)<-c("probe", "NCBI.gene.symbol", "EntrezGene.ID")
maux<-new("MAList", list(M=M, genes=genes))
#calling as function
object<-as(maux, "PAM50")
object
#same result with as.PAM50 function
object<-as.PAM50(maux)
object
```

##Example3: Work with PAM50 object: filtrate, classify and permutate-------

###1) Keep only annotated genes presents in PAM50 centroids

```r
object<-filtrate(object, verbose=TRUE)
```

###2) Get PAM50 subtypes without any normalization

```r
object<-classify(object, std="none", verbose=TRUE)
```

###3) Obtain the permutation subtype

Let’s run a quick example with 100 permutations. It is recommended at
pam50centroids

##least 10.000
object<-permutate(object, nPerm=100, pCutoff=0.01, corCutoff=0.1, keep=TRUE, seed=1234567890, verbose=TRUE)
object
##Now we can inspect the how the permutation went
head(permutation(object))
##Which parameters were used?
parameters(object)

##Example 4: Obtain summary statistics and reports--------------------------
##1) Let's check if we have a diagonal contingency matrix, i. e., no mistake
##is made in subtype assesment.
summary(object)

##2) Let's take a look at how the patient genes behave according
## to PAM50
subjectReport(object, subject=1)

##3) Just get a pdf with all the used subjects (PAM50 centroids in this
## example).
#databaseReport(object, fileName="PAM50.pdf", verbose=TRUE)

---

### Example PAM50 objects for pbcmc package

#### Description
The dataset corresponds to the Permutation-Based Confidence for Molecular Classification package PAM50 example objects, that was filtered, classified and permuted using the following parameters:

- **Permutations**: 10000
- **fdr**: 0.01
- **corCutoff**: 0.1
- **keep**: TRUE

#### Usage
```r
data(pam50centroids)
```

#### Format
pam50centroids corresponds with **pam50centroids** dataset available in genefu package.

#### Value
a PAM50 object with the results obtained for pam50centroids simulations under the given parameters (see Detail section.)

#### Author(s)
Cristobal Fresno <cfresno@bdmg.com.ar>, German A. Gonzalez <ggonzalez@bdmg.com.ar>, Andrea S. Llera <allera@leloir.org.ar> and Elmer Andres Fernandez <efernandez@bdmg.com.ar>
References


See Also

Other PAM50: as, classify,PAM50-method, filtrate,PAM50-method, permutate,PAM50-method, subjectReport,PAM50-method, subtypes,PAM50-method

---

**parameters**

Accessors for MolecularPermutationClassifier child class slots

**Description**

Slot setters/getters for MolecularPermutationClassifier hierarchy classes

**Usage**

```r
parameters(object)

## S4 method for signature 'MolecularPermutationClassifier'
parameters(object)

parameters(object) <- value

## S4 replacement method for signature 'MolecularPermutationClassifier'
parameters(object) <- value

## S4 method for signature 'MolecularPermutationClassifier'
exprs(object)

## S4 replacement method for signature 'MolecularPermutationClassifier,ANY'
exprs(object) <- value

## S4 method for signature 'MolecularPermutationClassifier'
annotation(object, ...)

## S4 replacement method for signature 'MolecularPermutationClassifier,ANY'
annotation(object) <- value

targets(object)

## S4 method for signature 'MolecularPermutationClassifier'
targets(object)
```

---

See Also

Other PAM50: as, classify,PAM50-method, filtrate,PAM50-method, permutate,PAM50-method, subjectReport,PAM50-method, subtypes,PAM50-method

---
targets(object) <- value
# S4 replacement method for signature 'MolecularPermutationClassifier'
targets(object) <- value

classification(object)
# S4 method for signature 'MolecularPermutationClassifier'
classification(object)

permutation(object)
# S4 method for signature 'MolecularPermutationClassifier'
permutation(object)

Arguments

object MolecularPermutationClassifier subclass object
value according to the function call:
  • parameters: named list with at least the following fields:
    $nPerm integer with number of permutations. Default: 1e4L
    $where character with significant value used. Default value is "fdr".
    $pCutoff numeric with p-value or fdr cutoff used, i.e., variable<pCutoff.
      Default: 0.01
    $corCutoff numeric with correlation difference between classes cutoff used,
      i.e., |ρ(profile, classA) − ρ(profile, classB)| > corCutoff
    $keep should null distribution simulation values be kept?. Default: FALSE
  • annotation: data.frame with individual annotations (genes, etc). Minimal
    compulsory fields are:
    $probe same characters as in row.names(M).
    $EntrezGene.ID integer with NCBI Entrez Data Base.
    $NCBI.gene.symbol character with gene mnemonic, a.k.a. gene symbol.
  • exprs: matrix with gene exprs profile, where genes are in rows and subjects
    as columns, a.k.a., M matrix.
  • targets: data.frame with additional subject data.
  ...
  additional parameters according to function call.

Value

according to function call one of the following objects:

parameters named list see value parameter
exprs matrix with gene exprs profile, where genes are in rows and subjects as columns,
      a.k.a., M matrix.
annotation data.frame see value parameter
classification named list with at least the following fields:

$class factor with with all possible class levels.
Gene expression-based classifiers, known as molecular signatures (MS), are a set of genes coordinately expressed and an algorithm that use these data to predict disease subtypes, response to therapy, disease risk or clinical outcome (Andre et al. 2006). They are especially important in breast cancer (BC) where several MS are currently on the market like PAM50 (Perou et al. 2000 & 2010), Prosigna [www.prosigna.com], Oncotype DX [www.oncotypedx.com], MammaPrint [www.agendia.com], etc. As far as the authors know, these classifiers do not give a real uncertainty of the classification at all. This package characterizes MS classification uncertainty. In order to
achieve this goal, synthetic simulated subjects are obtained by permutations of gene labels. Then, each synthetic subject is tested against the classifier corresponding subtype to build the null distribution, thus, classification confidence measurement can be provided for each subject. In this context, subjects belonging to the null distribution (random or noisy individuals) are not assigned (NA) to any class. On the contrary, if reliable results are obtained, subjects could be either assigned (A) to the more reliably subtype or marked as ambiguous (AMB) if proximal to two or more reliable subtypes. In the later, the combinations of classes are given. At present, it is only implemented for genefu’s PAM50 package (Haibe-Kains et al. 2014) but it can easily be extended to other MS. This package includes the following features:

- Implemented classifier:
  1. PAM50.

- Single subject classification:
  1. No pilot study needs to be carried out to obtain classification uncertainty.
  2. No normalization is required. If required, external database normalization, genefu normalization alternatives (scale/robust) or even gene median can be applied before simulations.

- Classification:
  1. The original PAM50 calls obtained by genefu.
  2. The proposed classification scheme: Assigned (PAM50 call), Not Assigned (NA) or Ambiguous (reliable PAM50 class combinations).
  3. Classification significance p-value or False Discovery Rate (FDR).
  4. Observed subject Spearman’s correlation for each breast cancer subtype.

- Physician treatment decision support:
  1. A friendly subject report is provided which includes summary data such as subtype centroid Spearman’s correlation, p-value and FDR for each subtype, original PAM50 classification and the recommended strategy (assigned, not assigned or ambiguous classes).
  2. Scatter plot of the observed gene-expression (subject) versus PAM50 centroids panel, plus the corresponding linear regression fit.
  3. Null distribution boxplot, plus observed (subject) value.

Author(s)

Cristobal Fresno <cfresno@bdmg.com.ar>, German A. Gonzalez <ggonzalez@bdmg.com.ar>, Andrea S. Llera <allera@leloir.org.ar> and Elmer Andres Fernandez <efernandez@bdmg.com.ar>

References

**Description**

Calculate the null Spearman’s ρ distribution of each subtype by means of gene label permutation, in order to evaluate if the observed values could be obtained by random change.

**Usage**

```r
## S4 method for signature 'PAM50'
permutate(object, nPerm = 10000, pCutoff = 0.01,
          where = "fdr", keep = FALSE, corCutoff = 0.1, seed = 1234567890,
          BPPARAM = bpparam(), verbose = getOption("verbose", default = TRUE))
```

**Arguments**

- `object` a MolecularPermutationClassifier subclass object.
- `nPerm` integer with number of permutations. Default: 1e4L
- `pCutoff` numeric with p-value or fdr cutoff used, i.e., variable<pCutoff. Default: 0.01
- `where` character with significant value used. Default value is “fdr”.
- `keep` should null distribution simulation values be kept?. Default: FALSE
- `corCutoff` numeric with correlation difference between classes cutoff used, i.e., \(|\rho(\text{profile}, \text{class}_A) - \rho(\text{profile}, \text{class}_B)| > corCutoff\). Default: 0.1
- `seed` integer to use as random seed. Default: 1234567890.
- `BPPARAM` an optional BiocParallelParam instance determining the parallel back-end to be used during evaluation, or a list of BiocParallelParam instances, to be applied in sequence for nested calls to bplapply. Default=bpparam().
- `verbose` should the user feedback be displayed? By default value is "verbose" global option parameter, if present, or FALSE otherwise.

**Value**

A PAM50 object with the following updated slots:

- `@permutation` $pvalues numeric matrix with subtype pvalues obtained as the number of times the permuted correlation is greater or equal the observed correlation divided the number of permutations.
- `@parameters` $nPerm, $pCutoff, $where and $keep updated accordingly.

**Author(s)**

Cristobal Fresno <cfresno@bdmg.com.ar>, German A. Gonzalez <ggonzalez@bdmg.com.ar>, Andrea S. Llera <allera@leloir.org.ar> and Elmer Andres Fernandez <efernandez@bdmg.com.ar>
References


See Also

PAM50 for a complete example.

Other PAM50: as, classify,PAM50-method, filtrate,PAM50-method, pam50centroids, subjectReport,PAM50-method, subtypes,PAM50-method

Examples

###Using pam50centroids package example data
data(pam50centroids)
pam50centroids
pam50centroids<-filtrate(pam50centroids, verbose=TRUE)
pam50centroids<-classify(pam50centroids, std="none", verbose=TRUE)

###Let's run a quick example with 100 permutations. It is recommended at least 10.000
pam50centroids<-permutate(pam50centroids, nPerm=100, pCutoff=0.01, corCutoff=0.1, verbose=TRUE)
pam50centroids

---

show Show a MolecularPermutationClassifier subclass object

Description

Basic MolecularPermutationClassifier class information display function (slots, dimensions, etc).

Usage

## S4 method for signature 'MolecularPermutationClassifier'
show(object)

Arguments

object an object of MolecularPermutationClassifier class hierarchy

Value

console messages displaying the class content
subjectReport,PAM50-method

Author(s)

Cristobal Fresno <cfresno@bdmg.com.ar>, German A. Gonzalez <ggonzalez@bdmg.com.ar>, Andrea S. Llera <allera@leloir.org.ar> and Elmer Andres Fernandez <efernandez@bdmg.com.ar>

See Also

PAM50 for a complete example.

Other MolecularPermutationClassifier: MolecularPermutationClassifier-class, parameters

Examples

##For an empty object
object<-PAM50()
object

##Using pam50centroids package example data
data(pam50centroids)
pam50centroids

subjectReport,PAM50-method

PAM50 permutation test results reports

Description

subjectReport is basically a grid.arrange object which basically consists of three main parts: a summary table, a two row ggplot2 facet_wrap with scatter ggplots (Wickham 2009) of subject expression and PAM50 centroids (Perou et al. 2000 & 2010) and a textGrob with the simulation parameter used. Particularly:

tableGrob with the following fields:

$Summary subject name, PAM50 and Permuted subtype

$Fields for the five PAM50 subtypes:

• Correlation: PAM50 centroid correlation with observed subject exprs.
• p-value: permutation p-value obtained using the simulation.
• FDR: adjusted p-value using False Discovery Rate.

ggplot facet_wrap two rows to display scatter subject exprs vs PAM50 centroids, in addition to a the linear regression fix. If subject, has an unique subtype, then the graph is in red. In addition, if simulated permutations were run with keep=TRUE option, then null distribution boxplots are plotted with observed correlations as a big round point.

textGrob the permutation @parameter slot used in the simulation.

Usage

## S4 method for signature 'PAM50'
subjectReport(object, subject)

## S4 method for signature 'PAM50'
databaseReport(object, fileName, ...,
subjectReport.PAM50-method

verbose = getOption("verbose", default = TRUE))

## S4 method for signature 'PAM50'
summary(object, ...)

Arguments

object   a PAM50 object.
subject   integer to select the appropriate subject to report.
fileName   character with the name of the pdf report file to save.
...   additional parameters for pdf function call.
verbose   should the user feedback be displayed? By default value is "verbose" global option parameter, if present, or FALSE otherwise.

Details

summary it basically prints descriptive data of PAM50 dataset, the test parameters used, a frequency table of PAM50 Subtypes and a contingency table with Classes vs PAM50 Subtypes.
databaseReport basically is a pdf report where the first page is a global summary of the database, i.e., a summary contingency table of permutation test classes against original PAM50 subtypes results. The following pages are the database respective subjectReport outputs.

Value

depending on function call:
subjectReport   a grid.arrange object.
databaseReport   a pdf file with database summary and subjectReports.
summary   Console summary statistics plus a data.frame

Author(s)

Cristobal Fresno <cfresno@bdmg.com.ar>, German A. Gonzalez <ggonzalez@bdmg.com.ar>, Andrea S. Llera <allera@leloir.org.ar> and Elmer Andres Fernandez <efernandez@bdmg.com.ar>

References


See Also

PAM50 for a complete example.

Other PAM50: as, classify, PAM50-method, filtrate, PAM50-method, pam50centroids, permutate, PAM50-method, subtypes, PAM50-method

Other PAM50: as, classify, PAM50-method, filtrate, PAM50-method, pam50centroids, permutate, PAM50-method, subtypes, PAM50-method

Other PAM50: as, classify, PAM50-method, filtrate, PAM50-method, pam50centroids, permutate, PAM50-method, subtypes, PAM50-method
Examples

```r
## Using pam50centroids package example data
data(pam50centroids)
pam50centroids

## This object has already run filtrate, classify and permutate. So, now
## we can obtain some reports:
## 1) database summary
summary(pam50centroids)

## 2) Individual subject report. If keep=FALSE boxplot panel is not available
subjectReport(pam50centroids, subject=1)  ## Basal subtype
subjectReport(pam50centroids, subject=1)  ## Her2 subtype

## 3) Complete database report
#databaseReport(pam50centroids, fileName="PAM50.pdf", verbose=TRUE)
```

Description

PAM50 subtypes are obtained using permuted test results. The idea is to give confidence in PAM50 subtype assessment (Perou et al. 2000 & 2010). In this context, the observed Spearman’s $\rho$ correlation is tested against the null distribution obtained for each subtype. Then, only significant correlations are used in according to the following scheme:

**Not assigned** all subtype have $\text{fdr} > \text{pcutoff}$. Hence, there is evidence that the observed $\rho$ can be obtained by random chance.

**Assigned** only one $\text{fdr} \leq \text{pcutoff}$. There is not enough evidence to say that the observed $\rho$ does not belong to the null distribution.

**Ambiguous** more than one have $\text{fdr} \leq \text{pcutoff}$. Then, one of the following alternatives holds given the result of $|\rho(\text{profile}, \text{class}_A) - \rho(\text{profile}, \text{class}_B)| > \text{corCutoff}$.

**Assigned** If the statement is TRUE.

**Ambiguous** If the statement is FALSE.

Under the above scheme, the physician has an objective measurement to support the patient treatment decision. Both, with the given permuted subtype and by interpreting the p-value or fdr of each subtype null distribution test.

Usage

```r
## S4 method for signature 'PAM50'
subtypes(object, pCutoff = 0.01, corCutoff = 0.1,
where = c("fdr", "pvalue")[1])
```

Arguments

- `object`: a `MolecularPermutationClassifier` subclass object.
- `pCutoff`: numeric with p-value/fdr cutoff used depending on "where" selection. Default: 0.01.
corCutoff numeric with correlation difference between classes cutoff used, i.e., \( |p(profile, class_A) - p(profile, class_B)| > corCutoff \). Default 0.1

where character with significant value used. Default value is "fdr".

Value

a PAM50 object with the updated slots:

@permutation $subtype data.frame with the following fields
$PAM50 the original PAM50 subtype
$Permuted factor with the following levels:
  • "Not assigned": all subtype have fdr > pcutoff
  • "Assigned": only one fdr <= pcutoff
  • "Ambiguous": more than one fdr <= pcutoff
$Classes a character according to "Permuted" field:
  • the unique PAM50 subtype if "Assigned"
  • a combination for "Ambiguous" or
  • NA if "Not assigned".
$Class idem as Classes but "Ambiguous" is set to PAM50 calls
$Subtype Classes but "Ambiguous" is kept as "Ambiguous" string.

@parameters SpCutoff, ScorCutoff and $where are updated accordingly.

Author(s)

Cristobal Fresno <cfresno@bdmg.com.ar>, German A. Gonzalez <ggonzalez@bdmg.com.ar>, Andrea S. Llera <allera@leloir.org.ar> and Elmer Andres Fernandez <efernandez@bdmg.com.ar>

References


See Also

PAM50 for a complete example.

Other PAM50: as, classify,PAM50-method, filtrate,PAM50-method, pam50centroids, permutate,PAM50-method, subjectReport,PAM50-method

Examples

```r
##Using pam50centroids package example data, which already had been
##filtrated, classified and permutated.
data(pam50centroids)
summary(pam50centroids)

##Now, let's change pCutoff and corCutoff without the need to run pemutate
##again
pam50centroids<-subtypes(pam50centroids, pCutoff=0.01, corCutoff=Inf,
where="fdr")
pam50centroids
summary(pam50centroids)#Note that only Basal is not Ambiguous
```

```
Index

*Topic Molecular
  pbcmcPackage, 17

*Topic PAM50
  pbcmcPackage, 17

*Topic Signature
  pbcmcPackage, 17

*Topic centroids
  pam50centroids, 14

*Topic datasets
  pam50centroids, 14

*Topic pam50
  pam50centroids, 14

annotation, 10, 11
annotation (parameters), 15
annotation,MolecularPermutationClassifier-method (parameters), 15
annotation<-(parameters), 15
annotation<-,MolecularPermutationClassifier,ANY-method (parameters), 15
as, 2, 2, 5, 7, 15, 20, 22, 24
as.PAM50, 10

classification, 10, 11
classification (parameters), 15
classification,MolecularPermutationClassifier-method (parameters), 15
classification-methods (parameters), 15
classify, 11, 14
classify (filtrate), 5
classify,PAM50-method, 4
classify-methods (filtrate), 5
databaseReport, 11
databaseReport (filtrate), 5
databaseReport,PAM50-method (subjectReport,PAM50-method), 21
databaseReport-methods (filtrate), 5

eprs, 10, 11
eprs (parameters), 15
eprs,MolecularPermutationClassifier-method (parameters), 15

exprs<-(parameters), 15
exprs<-,MolecularPermutationClassifier,ANY-method (parameters), 15
filtrate, 5, 9, 11, 13, 14
filtrate,PAM50-method, 7
filtrate-methods (filtrate), 5
loadBCDataset, 6, 8, 10, 11, 13
loadBCDataset, classGeneratorFunction-method (loadBCDataset), 8
loadBCDataset-methods (loadBCDataset), 8
MolecularPermutationClassifier-class, 9

PAM50, 3, 5–7, 9–11, 14, 17, 20–22, 24
PAM50 (PAM50-class), 11
PAM50-class, 11
PAM50-methods (parameters), 3, 5, 7, 14, 20, 22, 24
parameters, 10, 11, 15, 21
parameters,MolecularPermutationClassifier-method (parameters), 15
parameters-methods (parameters), 15
parameters<-(parameters), 15
parameters<-,MolecularPermutationClassifier-method (parameters), 15
parameters<--methods (parameters), 15
pbcmcPackage, 17
pbcmcPackage-package (pbcmcPackage), 17
permutate, 11, 14
permutate (filtrate), 5
permutate,PAM50-method, 19
permutate-methods (filtrate), 5
permutation, 10, 11
permutation (parameters), 15
permutation,MolecularPermutationClassifier-method (parameters), 15
permutation-methods (parameters), 15
print, 10

setAs, 10
show, 10, 11, 17, 20
show,MolecularPermutationClassifier-method (show), 20

25
subjectReport, 11
subjectReport (filtrate), 5
subjectReport,PAM50-method, 21
subjectReport-methods (filtrate), 5
subtypes, 11
subtypes (filtrate), 5
subtypes,PAM50-method, 23
subtypes-methods (filtrate), 5
summary, 10
summary,PAM50-method
(subjectReport,PAM50-method), 21

targets, 10, 11
targets (parameters), 15
targets,MolecularPermutationClassifier-method (parameters), 15
targets-methods (parameters), 15
targets<-(parameters), 15
targets<-,MolecularPermutationClassifier-method (parameters), 15
targets<--methods (parameters), 15