Package 'PCAPAM50'

January 9, 2025

Type Package

Title Enhanced 'PAM50' Subtyping of Breast Cancer

Version 1.0.3

Maintainer Praveen-Kumar Raj-Kumar <p.rajkumar@wriwindber.org>

Description Accurate classification of breast cancer tumors based on gene expression data is not a trivial task, and it lacks standard practices.The 'PAM50' classifier, which uses 50 gene centroid correlation distances to classify tumors, faces challenges with balancing estrogen receptor (ER) status and gene centering. The 'PCAPAM50' package leverages principal component analysis and iterative 'PAM50' calls to create a gene expression-based ER-balanced subset for gene centering, avoiding the use of protein expression-based ER data resulting into an enhanced Breast Cancer subtyping.

License GPL (>= 3)

Encoding UTF-8

Suggests testthat, knitr, rmarkdown

Config/testthat/edition 3

RoxygenNote 7.2.3

Imports Biobase, lattice, ComplexHeatmap, impute

VignetteBuilder knitr

NeedsCompilation no

Author Praveen-Kumar Raj-Kumar [aut, cre, cph],
Boyi Chen [aut],
Ming-Wen Hu [aut],
Tyler Hohenstein [aut],
Jianfang Liu [aut],
Craig D. Shriver [aut],
Xiaoying Lin [aut, cph],
Hai Hu [aut, cph]

Repository CRAN

Date/Publication 2025-01-09 17:10:05 UTC

Contents

Index

---

makeCalls.ihc   *Make a Conventional PAM50 Intrinsic Subtype Calls*

Description

This function processes clinical and preprocessed PAM50 expression data to form an estrogen receptor (ER)-balanced set based on IHC classification. The ER-balanced set is created by distinguishing between ER-negative and ER-positive cases, and it produces conventional PAM50 intrinsic subtype calls.

Usage

```
makeCalls.ihc(df.cln, seed=118, mat, outDir=NULL)
```

Arguments

df.cln	Data frame of clinical data; It should include the columns 'PatientID' and 'IHC'.
seed	Seed for random number generation to ensure reproducibility. Default is 118.
mat	Matrix of preprocessed PAM50 expression data.
outDir	Directory for output files.If NULL, a subdirectory named "Calls.PAM50" within the temporary directory will be used.

Value

Returns a list containing:

Int.sbs	Data frame with integrated subtype and clinical data.
score.fl	Data frame with scores from subtype predictions.
mdns.fl	Data frame with median values for each gene in the ER-balanced set.
SBS.colr	Colors associated with each subtype from the prediction results.
outList	Detailed results from subtype prediction functions.

See Also

prcomp, merge, set.seed

Examples

```
data_path <- system.file("extdata", "Sample_IHC_PAM_Mat.Rdat", package = "PCAPAM50")
load(data_path) # Loads Test.ihc and Test.matrix


# Prepare the data
Test.ihc$ER_status <- rep("NA", length(Test.ihc$PatientID))
Test.ihc$ER_status[grep("^L",Test.ihc$IHC)] = "pos"
Test.ihc$ER_status[-grep("^L",Test.ihc$IHC)] = "neg"
Test.ihc <- Test.ihc[order(Test.ihc$ER_status, decreasing = TRUE),]
Test.matrix <- Test.matrix[, Test.ihc$PatientID]


df.cln <- data.frame(PatientID = Test.ihc$PatientID, IHC = Test.ihc$IHC, stringsAsFactors = FALSE)




# Call the function
result <- makeCalls.ihc(df.cln=df.cln, seed = 118, mat = Test.matrix, outDir=NULL)
```

---

makeCalls.PC1ihc	*Make intermediate intrinsic subtype calls*

Description

This function processes clinical IHC subtyping data and preprocessed PAM50 gene expression data to form a gene expression-guided ER-balanced set.This set is created by combining IHC classification information and using principal component 1 (PC1) to guide the separation.The function computes the median for each gene in this ER-balanced set, updates a calibration file, and runs subtype prediction algorithms to generate intermediate intrinsic subtype calls based on the PAM50 method. Various diagnostics and subtyping results are returned.

Usage

```
makeCalls.PC1ihc(df.cln, seed = 118, mat, outDir=NULL)
```

Arguments

df.cln	Data frame of clinical data; It should include the columns 'PatientID' and 'IHC'.
seed	Seed for random number generation to ensure reproducibility. Default is 118.
mat	Matrix of preprocessed PAM50 expression data.
outDir	Directory for output files.If NULL, a subdirectory named "Calls.PC1ihc" within the temporary directory will be used.

Value

Returns a list containing:

`Int.sbs`	Data frame with integrated subtype and clinical data.
`score.fl`	Data frame with scores from subtype predictions.
`mdns.fl`	Data frame with median values for each gene in the ER-balanced set.
`SBS.colr`	Colors associated with each subtype from the prediction results.
`outList`	Detailed results from subtype prediction functions.
`PC1cutoff`	Cutoff values for PC1 used in subsetting.
`DF.PC1`	Data frame of initial PCA results merged with clinical data.

See Also

prcomp, merge, set.seed

Examples

```
data_path <- system.file("extdata", "Sample_IHC_PAM_Mat.Rdat", package = "PCAPAM50")
load(data_path) # Loads Test.ihc and Test.matrix


# Prepare the data
Test.ihc$ER_status <- rep("NA", length(Test.ihc$PatientID))
Test.ihc$ER_status[grep("^L",Test.ihc$IHC)] = "pos"
Test.ihc$ER_status[-grep("^L",Test.ihc$IHC)] = "neg"
Test.ihc <- Test.ihc[order(Test.ihc$ER_status, decreasing = TRUE),]
Test.matrix <- Test.matrix[, Test.ihc$PatientID]


df.cln <- data.frame(PatientID = Test.ihc$PatientID, IHC = Test.ihc$IHC, stringsAsFactors = FALSE)



# Call the function
result <- makeCalls.PC1ihc(df.cln=df.cln, seed = 118, mat = Test.matrix, outDir=NULL)
```

---

makeCalls.v1PAM *Make PCAPAM50 calls*

Description

This function uses the intermediate intrinsic subtype calls and preprocessed PAM50 gene expression data to create an ER-balanced set and produces PCAPAM50 Calls.

Usage

```
makeCalls.v1PAM(df.pam, seed = 118, mat, outDir=NULL)
```

Arguments

`df.pam`	Data frame of PAM data; It should include the columns 'PatientID' and 'PAM50'.
`seed`	Seed for random number generation to ensure reproducibility.
`mat`	Matrix of preprocessed PAM50 expression data.
`outDir`	Directory for output files.If NULL, a subdirectory named "Calls.PCAPAM50" within the temporary directory will be used.

Value

Returns a list containing:

`Int.sbs`	Data frame with integrated subtype and clinical data.
`score.fl`	Data frame with scores from subtype predictions.
`mdns.fl`	Data frame with median values for each gene in the ER-balanced set.
`SBS.colr`	Colors associated with each subtype from the prediction results.
`outList`	Detailed results from subtype prediction functions.

See Also

[prcomp](), [merge](), [set.seed]()

Examples

```
data_path <- system.file("extdata", "Sample_IHC_PAM_Mat.Rdat", package = "PCAPAM50")
load(data_path) # Loads Test.ihc and Test.matrix


# Prepare the data
Test.ihc$ER_status <- rep("NA", length(Test.ihc$PatientID))
Test.ihc$ER_status[grep("^L",Test.ihc$IHC)] = "pos"
Test.ihc$ER_status[-grep("^L",Test.ihc$IHC)] = "neg"
Test.ihc <- Test.ihc[order(Test.ihc$ER_status, decreasing = TRUE),]
Test.matrix <- Test.matrix[, Test.ihc$PatientID]

df.cln <- data.frame(PatientID = Test.ihc$PatientID, IHC = Test.ihc$IHC, stringsAsFactors = FALSE)
outDir <- "Call.PC1"

# Make a secondary ER-balanced subset and derive intermediate intrinsic subtype calls
result <- makeCalls.PC1ihc(df.cln=df.cln, seed = 118, mat = Test.matrix, outDir=outDir)

df.pc1pam = data.frame(PatientID=result$Int.sbs$PatientID,
  PAM50=result$Int.sbs$Int.SBS.Mdns.PC1ihc,
IHC=result$Int.sbs$IHC,
stringsAsFactors=FALSE) ### IHC column is optional
```

```
# Make a tertiary ER-balanced set and PCAPAM50 calls
res <- makeCalls.v1PAM(df.pam = df.pc1pam, seed = 118, mat = Test.matrix, outDir=NULL)
```

---

my.plotPCA   *Modeling after plotPCA of DESeq*

Description

Modeling after plotPCA of DESeq

Usage

```
my.plotPCA(x, intgroup, ablne = 0,
          colours = c("red","hotpink","darkblue", "lightblue","red3","hotpink3",
            "royalblue3","lightskyblue3"),
            LINE.V = TRUE)
```

Arguments

x	An ExpressionSet object, with matrix data (x) in 'assay(x)', produced for example by ExpressionSet(assayData=Test.matrix, phenoData=phenoData)
intgroup	Subtype condition: a character vector of names in 'colData(x)' to use for grouping.
ablne	An x-axis coordinate for drawing a vertical line. Default is 0.
colours	Colors for subtypes present in the condition.
LINE.V	Determines whether or not to draw line. Default is TRUE.

Value

Returns an image containing:

pcafig	The plot.

See Also

prcomp, merge, set.seed

Examples

```
library("Biobase")

data_path <- system.file("extdata", "Sample_IHC_PAM_Mat.Rdat", package = "PCAPAM50")
load(data_path) # Loads Test.ihc and Test.matrix

pData = data.frame(condition=Test.ihc$IHC)
rownames(pData) = Test.ihc$PatientID
phenoData = new("AnnotatedDataFrame", data=pData)#, varMetadata=Metadata
XSet      = ExpressionSet(assayData=Test.matrix, phenoData=phenoData)
my.plotPCA(XSet, intgroup=pData$condition, ablne=2.4,
 colours = c("hotpink","darkblue","lightblue","lightblue3","red"),
 LINE.V = TRUE)
```

Index