## Package ‘TNBC.CMS’

April 2, 2024

**Type** Package

**Date** 2021-11-21

**Title** TNBC.CMS: Prediction of TNBC Consensus Molecular Subtypes

**Version** 1.18.0

**Author** Doyeong Yu, Jihyun Kim, In Hae Park, Charny Park

**Maintainer** Doyeong Yu <parklab.bi@gmail.com>

**Description** This package implements a machine learning-based classifier for the assignment of consensus molecular subtypes to TNBC samples. It also provides functions to summarize genomic and clinical characteristics.

**License** GPL-3

**Encoding** UTF-8

**biocViews** Classification, Clustering, GeneExpression, GenePrediction, SupportVectorMachine

**RoxygenNote** 6.1.1

**Depends** R (>= 3.6.0), e1071, quadprog, SummarizedExperiment

**Imports** GSVA (>= 1.26.0), pheatmap, grDevices, RColorBrewer, pracma, GGally, R.utils, forestplot, ggplot2, ggrepur, survival, grid, stats, methods

**Suggests** knitr

**VignetteBuilder** knitr

**git url** https://git.bioconductor.org/packages/TNBC.CMS

**git branch** RELEASE_3_18

**git last commit** 7ce7eb8

**git last commit date** 2023-10-24

**Repository** Bioconductor 3.18

**Date/Publication** 2024-04-01
computeDS

R topics documented:

computeDS ................................................................. 2
computeESTIMATEscore .................................................. 3
computeGES ................................................................. 4
computexCellScore ....................................................... 5
GSE25055 ................................................................. 6
performGSVA ............................................................... 6
plotHR ................................................................. 7
plotKM ................................................................. 8
predictCMS ............................................................... 9

Description

Computes drug signature scores. Also draws heatmap representing the average signature scores for each subtype.

Usage

computeDS(expr, pred, gene.set = NULL)

Arguments

expr A SummarizedExperiment object or a matrix containing gene expression profiles. If input is a SummarizedExperiment, the first element in the assays list should be a matrix of gene expression. Rows and columns of the gene expression matrix correspond to genes and samples, respectively (rownames must be to gene symbols).

pred A vector of predicted consensus molecular subtypes.

gene.set A user-provided list of gene sets associated with drug response. Names of gene sets must follow the format of [DRUG NAME]_[RESISTANCE/RESPONSE]_[UP/DN] (e.g. CISPLATIN_RESISTANCE_DN).

Details

Drug signature scores are the average of expression values of genes included in gene sets from MSigDB.

Value

A matrix of drug signature scores.
computeESTIMATEscore 3

References


Examples

# Load gene expression profiles of TNBC samples
data(GSE25055)

# Predict consensus molecular subtypes of TNBC samples
prediction <- predictCMS(expr = GSE25055)

# Compute drug signature scores
resultDS <- computeDS(expr = GSE25055, pred = prediction)

computeESTIMATEscore  Computation of stromal and immune scores

Description

Computes stromal and immune scores. This function was borrowed from the estimate package and changed to accept R object as input.

Usage

computeESTIMATEscore(mat)

Arguments

mat  A matrix of gene expression with genes in rows and samples in columns (row-names corresopnding to gene symbol).

Value

A data frame containing stromal and immune scores
computeGES

Computation of gene expression signature scores.

Description

Computes gene expression signature scores. Also draws boxplots representing the average signature scores for each subtype.

Usage

computeGES(expr, pred, rnaseq = FALSE)

Arguments

expr A SummarizedExperiment object or a matrix containing gene expression profiles. If input is a SummarizedExperiment, the first element in the assays list should be a matrix of gene expression. Rows and columns of the gene expression matrix correspond to genes and samples, respectively (rownames must be to gene symbols).

pred A vector of predicted consensus molecular subtypes.

rnaseq logical to determine if input data is RNA-Seq gene expression profile. By default, it is FALSE.

Details

computeGES calculates the following 7 gene expression signature scores:

- EMT (epithelial-mesenchymal transition): average of expression values of genes included in the EMT signature published by Tan et al. (2014).
- Stromal: stromal score representing the presence of stromal cells in tumor tissues (computed using the ESTIMATE algorithm).
- Immune: immune score representing the presence of immune cells in tumor tissues (computed using the ESTIMATE algorithm).
- Microenvironment: microenvironment score representing the sum of all immune and stromal cell types (computed using xCell).
- Stemness: stemness index computed using the method developed by Malta et al. (2018).
- Hormone: average of expression values of AR, ERBB2, ESR1, and PGR.
- CIN (chromosomal instability): average of expression values of genes included in the CIN70 signature published by Carter et al. (2006).

Value

A matrix of gene expression signature scores.
References


Examples

```r
# Load gene expression profiles of TNBC samples
data(GSE25055)

# Predict consensus molecular subtypes of TNBC samples
prediction <- predictCMS(expr = GSE25055)

# Compute gene expression signature scores
resultGES <- computeGES(expr = GSE25055, pred = prediction, rnaseq = FALSE)
```

computexCellScore  Computation of microenvironment score

Description

Computes a microenvironment score. This function wraps around the xCellAnalysis function of the xCell package to compute a microenvironment score.

Usage

```r
computexCellScore(mat, rnaseq)
```

Arguments

- **mat**: A matrix of gene expression with genes in rows and samples in columns (row-names corresopnding to gene symbol).
- **rnaseq**: logical to determine if input data is RNA-Seq gene expression profile

Value

A data frame containing stromal and immune scores
PerformGSVAData

Description

This is a TNBC microarray dataset from GSE25055 contained in a SummarizedExperiment object. It includes gene expression profiles and clinical information which can be accessed by the assays and colData functions, respectively. We obtained gene expression profiles of breast cancer samples from the curatedBreastData package and extracted TNBC samples based on the expression profiles and immunohistochemistry results.

Source


References


Examples

data(GSE25055)

#Access gene expression profiles
head(assays(GSE25055)[[1]])

#Access clinical information
head(colData(GSE25055))

---

performGSVA

Description

Performs GSVA on gene sets. Also draws a heatmap representing GSVA scores.

Usage

performGSVA(expr, pred, gene.set = NULL, gsva.kcdf = "Gaussian")
plotHR

Arguments

expr A SummarizedExperiment object or a matrix containing gene expression profiles. If input is a SummarizedExperiment, the first element in the assays list should be a matrix of gene expression. Rows and columns of the gene expression matrix correspond to genes and samples, respectively (rownames must be gene symbols).

pred A vector of predicted consensus molecular subtypes.

gene.set Gene sets provided as a list. If NULL, the hallmark pathway gene sets are used.

gsva.kcdf Kernel to be used in the estimation of the cumulative distribution function. By default, this is set to “Gaussian” which is suitable for continuous expression values. If expression values are counts, “Poisson” is recommended.

Details

This is a wrapper function of the gsva function in the GSVA package to compute GSVA enrichment scores per sample and produce a heatmap comparing them across consensus molecular subtypes.

Value

A matrix of GSVA enrichment scores.

References


Examples

# Load gene expression profiles of TNBC samples
data(GSE25055)

# Predict consensus molecular subtypes of TNBC samples
prediction <- predictCMS(expr = GSE25055)

# Perform GSVA on the hallmark pathway gene sets
resultGSVA <- performGSVA(expr = GSE25055, pred = prediction)

plotHR(expr, gene.symbol, pred, time, event, by.subtype = TRUE)
Arguments

expr  A SummarizedExperiment object or a matrix containing gene expression profiles. If input is a SummarizedExperiment, the first element in the assays list should be a matrix of gene expression. Rows and columns of the gene expression matrix correspond to genes and samples, respectively (rownames must be to gene symbols).

gene.symbol  A vector of gene symbols for which hazard ratios are computed.
pred  A vector of predicted consensus molecular subtypes.
time  A vector of the follow-up time.
event  A vector representing survival status (0 = alive, 1 = dead).
by.subtype  A logical to determine if subtype-specific hazard ratios are computed (default is TRUE).

Value

A forest plot of hazard ratios.

Examples

# Load gene expression profiles and clinical information of TNBC samples
data(GSE25055)
DFS.status <- colData(GSE25055)$DFS.status
DFS.month <- colData(GSE25055)$DFS.month

# Predict consensus molecular subtypes of TNBC samples
prediction <- predictCMS(expr = GSE25055)

# Forest plot of hazard ratios for input genes
plotHR(expr = GSE25055, gene.symbol = c("RECK", "RELN", "EHD4", "PRRX2"),
pred = prediction, time = DFS.month, event = DFS.status,
by.subtype = FALSE)

# Subtype-specific forest plot of hazard ratios for input genes
plotHR(expr = GSE25055, gene.symbol = c("RECK", "RELN", "EHD4", "PRRX2"),
pred = prediction, time = DFS.month, event = DFS.status,
by.subtype = TRUE)

plotKM  Subtype-specific survival curves

Description

Produces Kaplan-Meier survival curves for each subtype.

Usage

plotKM(pred, time, event)
predictCMS

Arguments

pred  A vector of predicted consensus molecular subtypes.
time  A vector of the follow-up time.
event A vector representing survival status (0 = alive, 1 = dead).

Value

A ggplot object.

Examples

# Load clinical information of TNBC samples
data(GSE25055)
DFS.status <- colData(GSE25055)$DFS.status
DFS.month <- colData(GSE25055)$DFS.month

# Predict consensus molecular subtypes of TNBC samples
prediction <- predictCMS(expr = GSE25055)

# Plot Kaplan-Meier curves for each subtype
plotKM(pred = prediction, time = DFS.month, event = DFS.status)

predictCMS  
TNBC consensus molecular subtype prediction

Description

Predicts the TNBC consensus molecular subtype of TNBC samples.

Usage

predictCMS(expr)

Arguments

expr A SummarizedExperiment object or a matrix containing gene expression profiles. If input is a SummarizedExperiment, the first element in the assays list should be a matrix of gene expression. Rows and columns of the gene expression matrix correspond to genes and samples, respectively (rownames must be to gene symbols).

Value

A vector of assigned subtypes.
Examples

# Load gene expression profiles of TNBC samples
data(GSE25055)

# Predict consensus molecular subtypes of TNBC samples
prediction <- predictCMS(expr = GSE25055)
table(prediction)
Index

* datasets
  GSE25055, 6

* internal
  computeESTIMATEscore, 3
  computexCellScore, 5

computeDS, 2
computeESTIMATEscore, 3
computeGES, 4
computexCellScore, 5

GSE25055, 6

performGSVA, 6
plotHR, 7
plotKM, 8
predictCMS, 9