Package ‘TNBC.CMS’

March 28, 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, ggpubr, 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 2024-03-27
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

References


Examples

```r
# 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

```r
computeESTIMATEscore(mat)
```

Arguments

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

Value

A data frame containing stromal and immune scores
### Description

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

### Usage

```r
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

`computexCellScore(mat, rnaseq)`

Arguments

- `mat`  
  A matrix of gene expression with genes in rows and samples in columns (row-names correspond 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
GSE25055

Example TNBC microarray data

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

Gene set variation analysis

Description

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

Usage

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

**Forest plot of hazard ratios**

**Description**

Produces a forest plot of hazard ratios for each gene. Also draws a forest plot of subtype-specific hazard ratios.

**Usage**

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)
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