Package 'TNBC.CMS'

April 12, 2022

Type Package

Date 2020-03-27

Title TNBC.CMS: Prediction of TNBC Consensus Molecular Subtypes

Version 1.10.0

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_14

git_last_commit dd4cd1f

git_last_commit_date 2021-10-26

Date/Publication 2022-04-12

R topics documented:

`computeDS` *Computation of drug signature scores.*

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 containig 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.

References

Liberzon, A. et al. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, 27, 1739-40.

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)
```

`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 containig 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

Aran, D. et al. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome biology*, 18, 220.

Carter, S.L. et al. (2006). A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nature genetics*, 38, 1043.

Malta, T.M. et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell*, 173, 338-354.

Tan, T.Z. et al. (2014). Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO molecular medicine*, 6, 1279-93.

Yoshihara, K. et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications*, 4, 2612.

Examples

```
# 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)
```

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

https://bioconductor.org/packages/release/data/experiment/html/curatedBreastData.html

References

Hatzis, C. et al. (2011). A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer *JAMA*, 305, 1873-81.

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")
```

Arguments

`expr`	A `SummarizedExperiment` object or a matrix containig 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`	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

Liberzon, A. et al. (2015). The molecular signatures database hallmark gene set collection. *Cell systems*, 1, 417-425.

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 *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 containig 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)
```

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 containig 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