Package 'CTexploreR'

May 19, 2024

Title Explores Cancer Testis Genes

Version 1.0.0

Description The CTexploreR package re-defines the list of Cancer Testis/Germline (CT) genes. It is based on publicly available RNAseq databases (GTEx, CCLE and TCGA) and summarises CT genes' main characteristics. Several visualisation functions allow to explore their expression in different types of tissues and cancer cells, or to inspect the methylation status of their promoters in normal tissues.

License Artistic-2.0

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

Depends R (>= 4.3), CTdata

Imports BiocGenerics, ComplexHeatmap, grid, SummarizedExperiment, GenomicRanges, IRanges, dplyr, tidyr, tibble, ggplot2, rlang, grDevices, stats, circlize, ggrepel, SingleCellExperiment, MatrixGenerics

Suggests BiocStyle, knitr, rmarkdown, testthat (>= 3.0.0)

biocViews Transcriptomics, Epigenetics, DifferentialExpression, GeneExpression, DNAMethylation, ExperimentHubSoftware, DataImport

VignetteBuilder knitr

URL https://github.com/UCLouvain-CBIO/CTexploreR

BugReports https://github.com/UCLouvain-CBIO/CTexploreR/issues

Config/testthat/edition 3

git_url https://git.bioconductor.org/packages/CTexploreR

git_branch RELEASE_3_19

git_last_commit 7af29e4

git_last_commit_date 2024-04-30

Repository Bioconductor 3.19

Date/Publication 2024-05-19

Author Axelle Loriot [aut, cre] (<https://orcid.org/0000-0002-5288-8561>),
Julie Devis [aut] (<https://orcid.org/0000-0001-5525-5666>),
Anna Diacofotaki [ctb],
Charles De Smet [ths],
Laurent Gatto [aut, ths] (<https://orcid.org/0000-0002-1520-2268>)

Maintainer Axelle Loriot <axelle.loriot@uclouvain.be>

Contents

`CCLE_expression` *Gene expression in CCLE Tumors*

Description

Plots an expression heatmap of genes in CCLE tumor cell lines.

Usage

```
CCLE_expression(
  genes = NULL,
  type = NULL,
  units = c("TPM", "log_TPM"),
  values_only = FALSE
)
```

Arguments

genes	character nameing the selected genes. The default value, NULL, takes all CT genes.
type	character() describing the tumor cell line(s) type to be plotted. Allowed cell lines are "Ovarian", "Leukemia", "Colorectal", "Skin", "Lung", "Bladder", "Kidney", "Breast", "Pancreatic", "Myeloma", "Brain", "Sarcoma", "Lymphoma", "Bone", "Neuroblastoma", "Gastric", "Uterine", "Head_and_Neck", "Bile_Duct" and "Esophageal".
units	character(1) with expression values unit. Can be "TPM" (default) or "log_TPM" (log(TPM + 1))
values_only	logical(1). If TRUE, values are returned instead of the heatmap (FALSE by default).

Value

A heatmap of selected genes in CCLE cell lines from specified type. If values_only is TRUE, expression values are returned instead.

Examples

```
## Not run:
CCLE_expression(
    genes = c("MAGEA1", "MAGEA3", "MAGEA4", "MAGEA6", "MAGEA10"),
    type = c("Skin", "Lung"), units = "log_TPM")

## End(Not run)
```

---

check_names *Check spelling of entered variables*

Description

Checks the spelling of a vector of entered variable(s) comparing it to a vector of valid names, and removes the ones that are absent from the vector of valid names.

Usage

```
check_names(variable, valid_vector)
```

Arguments

variable	character() containing the names of variables to check.
valid_vector	character() with valid variable names.

Value

A character with valid variables.

Examples

```
CTexploreR:::check_names(
    variable = c("Ovarian", "leukemia", "wrong_name"),
    valid_vector = c("ovarian", "leukemia")
)
```

`CT_correlated_genes` *Gene correlations in CCLE cancer cell lines*

Description

A function that uses expression data from CCLE cell lines and highlights genes correlated (or anti-correlated) with specified CT gene. Genes with a correlation coefficient above threshold are colored in red if they are CT genes or in blue, if not.

Usage

```
CT_correlated_genes(gene, corr_thr = 0.5, values_only = FALSE)
```

Arguments

`gene`	CT gene selected
`corr_thr`	`numeric(1)` with default 0.5. Genes with an absolute correlation coefficient (Pearson) higher than this threshold will be highlighted.
`values_only`	`logical(1)`, `FALSE` by default. If `TRUE`, the function will return the correlation coefficients with all genes instead of the plot.

Value

A plot where each dots represent the correlation coefficients (Pearson) between genes and the specified CT gene (entered as input). Genes with a correlation coefficient above threshold are colored in red if they are CT genes or in blue, if not. If `values_only = TRUE`, all correlations coefficients are returned instead.

Examples

```
## Not run:
CT_correlated_genes(gene = "MAGEA3")

## End(Not run)
```

CT_genes *CT genes description table*

Description

Cancer-Testis (CT) genes description, imported from CTdata

Usage

```
CT_genes
```

Format

A tibble object with 298 rows and 36 columns.

- Rows correspond to CT genes
- Columns give CT genes characteristics

Details

See CTdata::CT_genes documentation for details

Value

A tibble of all 298 CT genes with their characteristics

Source

See scripts/make_CT_genes.R in CTdata for details on how this list of curated CT genes was created.

Examples

```
CT_genes
```

DAC_induction *Gene expression in cells treated or not by a demethylating agent*

Description

Plots a heatmap of normalised gene counts (log-transformed) in a selection of cells treated or not by 5-Aza-2'-Deoxycytidine (DAC), a demethylating agent.

Usage

```
DAC_induction(genes = NULL, multimapping = TRUE, values_only = FALSE)
```

Arguments

genes	character nameing the selected genes. The default value, NULL, takes all CT genes.
multimapping	logical(1) defining whether to use multi-mapped gene expression dataset CTdata::DAC_treated_cell or DAC_treated_cells. Default is TRUE.
values_only	logical(1). If TRUE, the function will return the gene normalised logcounts in all samples instead of the heatmap. Default is FALSE.

Details

RNAseq data from cells treated or not with 5-aza downloaded from SRA. (SRA references and information about cell lines and DAC treatment are stored the colData of DAC_treated_cells). Data was processed using a standard RNAseq pipeline. hisat2 was used to align reads to grch38 genome. featurecounts was used to assign reads to genes. Note that -M parameter was used or not to allow or not counting multi-mapping reads.

Value

A heatmap of selected genes in cells treated or not by a demethylating agent. If values_only is TRUE, gene normalised logcounts are returned instead.

Examples

```
DAC_induction(genes = c("MAGEA1", "MAGEA3", "MAGEA4", "MAGEA6", "CTAG1A"))
DAC_induction(genes = c("MAGEA1", "MAGEA3", "MAGEA4", "MAGEA6", "CTAG1A",
    multimapping = FALSE))
```

GTEX_expression *Gene expression in normal tissues (GTEx)*

Description

Plots an expression heatmap of genes in normal tissues (GTEx database).

Usage

```
GTEX_expression(genes = NULL, units = c("TPM", "log_TPM"), values_only = FALSE)
```

Arguments

genes	`character` nameing the selected genes. The default value, `NULL`, takes all CT genes.
units	`character(1)` with expression values unit. Can be `"TPM"` (default) or `"log_TPM"` (log(TPM + 1)).
values_only	`logical(1)`. If `TRUE`, the function will return the expression values in all samples instead of the heatmap. Default is `FALSE`.

Value

A heatmap of selected genes expression in normal tissues. If `values_only = TRUE`, expression values are returned instead.

Examples

```
GTEX_expression(units = "log_TPM")
GTEX_expression(genes = c("MAGEA1", "MAGEA3"), units = "log_TPM")
```

HPA_cell_type_expression *Gene expression in different human cell types*

Description

Plots a heatmap of genes expression in the different human cell types based on scRNAseq data obtained from the Human Protein Atlas (https://www.proteinatlas.org)

Usage

```
HPA_cell_type_expression(
  genes = NULL,
  units = c("scaled", "TPM", "log_TPM"),
  scale_lims = NULL,
  values_only = FALSE
)
```

Arguments

genes	character nameing the selected genes. The default value, NULL, takes all CT genes.
units	character(1) with expression values unit. Can be "TPM", "log_TPM" (log(TPM + 1)) or "scaled" (scaled TPM values). Default is "scaled".
scale_lims	vector of length 2 setting the lower and upper limits of the heatmap colorbar.
values_only	logical(1). If TRUE, the function will return the SummarizedExperiment instead of the heatmap. Default is FALSE.

Value

A heatmap of selected CT genes expression in different human cell types. If values_only = TRUE, a SummarizedExperiment instead of the heatmap is returned instead.

Examples

```
HPA_cell_type_expression(
    genes = NULL, units = "scaled", scale_lims = NULL,
    values_only = FALSE)
HPA_cell_type_expression(
    genes = c("MAGEA1", "MAGEA3", "MAGEA4"),
    units = "TPM", scale_lims = c(0, 50),
    values_only = FALSE)
```

---

normal_tissues_mean_methylation

Promoter methylation of Cancer-Testis genes in normal tissues

---

Description

Plots a heatmap of mean promoter methylation levels of Cancer-Testis (CT) genes in normal tissues. Methylation levels in tissues correspond to the mean methylation of CpGs located in range of 1000 pb upstream and 200 pb downstream from gene TSS.

Usage

```
normal_tissues_mean_methylation(
  genes = NULL,
  values_only = FALSE,
  na.omit = TRUE
)
```

Arguments

genes	character nameing the selected genes. The default value, NULL, takes all CT genes.
values_only	logical(1), FALSE by default. If TRUE, the function will return the methylation values in all samples instead of the heatmap.
na.omit	logical(1) specifying if genes with missing methylation values in some tissues should be removed (TRUE by default). Note that no gene clustering will be done when methylation values are missing.

Value

Heatmap of mean promoter methylation of Cancer-Testis (CT) genes in normal tissues. If values_only = TRUE, methylation values are returned instead.

Examples

```
normal_tissues_mean_methylation()
normal_tissues_mean_methylation(c("MAGEA1", "MAGEA2", "MAGEA3", "MAGEA4"))
normal_tissues_mean_methylation(c("MAGEA1", "MAGEA2", "MAGEA3", "MAGEA4"),
    na.omit = FALSE)
```

---

normal_tissues_methylation

Methylation of CpGs located in Cancer-Testis promoters in normal tissues

---

Description

Plots a heatmap of the methylation of CpGs located in a Cancer-Testis (CT) promoter, in normal tissues. X-axis corresponds to the CpGs position (related to TSS).

Usage

```
normal_tissues_methylation(
  gene,
  nt_up = 1000,
  nt_down = 200,
  values_only = FALSE
)
```

Arguments

gene	Name of selected CT gene
nt_up	Number of nucleotides upstream the TSS to analyse (by default 1000, maximum value 5000)
nt_down	Number of nucleotides downstream the TSS to analyse (by default 200, maximum value 5000)

values_only Boolean (FALSE by default). If set to TRUE, the function will return the methylation values of all cytosines in the promoter instead of the heatmap.

Value

Heatmap of the methylation of CpGs located in a Cancer-Testis (CT) promoter, in normal tissues. If values_only = TRUE, methylation values are returned instead.

Examples

```
normal_tissues_methylation(gene = "TDRD1", 1000, 0)
```

---

normal_tissue_expression_multimapping

Expression values (TPM) of genes in normal tissues with or without multimapping

---

Description

Plots a heatmap of gene expression values in a set of normal tissues. Expression values (in TPM) have been evaluated by either counting or discarding multi-mapped reads. Indeed, many CT genes belong to gene families from which members have identical or nearly identical sequences. Some CT can only be detected in RNAseq data in which multimapping reads are not discarded.

Usage

```
normal_tissue_expression_multimapping(
  genes = NULL,
  multimapping = TRUE,
  units = c("TPM", "log_TPM"),
  values_only = FALSE
)
```

Arguments

genes character nameing the selected genes. The default value, NULL, takes all CT genes.

multimapping logical(1) that specifies if returned expression values must take into account or not multi-mapped reads. TRUE by default.

units character(1) with expression values unit. Can be "TPM" (default) or "log_TPM" (log(TPM + 1)).

values_only logical(1). If TRUE, the function will return the expression values in all samples instead of the heatmap. Default is FALSE.

Details

RNAseq data from a set of normal tissues were downloaded from Encode. (see inst/scripts/make_CT_normal_tissues_multim for fastq references) Fastq files were processed using a standard RNAseq pipeline including FastQC for the quality control of the raw data, and trimmomatic to remove low quality reads and trim the adapter from the sequences. hisat2 was used to align reads to grch38 genome. featurecounts was used to assign reads to genes using Homo_sapiens.GRCh38.105.gtf.

Two different pipelines were run in order to remove or not multi-mapping reads. When multimapping was allowed, hisat2 was run with -k 20 parameter (reports up to 20 alignments per read), and featurecounts was run with -M parameter (multi-mapping reads are counted).

Value

A heatmap of selected gene expression values in a set of normal tissues calculated by counting or discarding multi-mapped reads. If `values_only = TRUE`, gene expression values are returned instead.

Examples

```
normal_tissue_expression_multimapping(
    genes = c("GAGE13", "CT45A6", "NXF2", "SSX2", "CTAG1A",
    "MAGEA3", "MAGEA6"), multimapping = FALSE)
normal_tissue_expression_multimapping(
    genes = c("GAGE13", "CT45A6", "NXF2", "SSX2", "CTAG1A",
    "MAGEA3", "MAGEA6"), multimapping = TRUE)
```

---

prepare_TCGA_methylation_expression

Prepare methylation and expression data of a gene in TCGA tumors

---

Description

Creates a Dataframe giving for each TCGA sample, the methylation level of a gene (mean methylation of probes located in its promoter) and the expression level of the gene (TPM value).

Usage

```
prepare_TCGA_methylation_expression(
  tumor = "all",
  gene = NULL,
  nt_up = NULL,
  nt_down = NULL,
  include_normal_tissues = FALSE
)
```

Arguments

tumor	character defining the TCGA tumor type. Can be one of "SKCM", "LUAD", "LUSC", "COAD", "ESCA", "BRCA", "HNSC", or "all" (default).
gene	character selected CT gene.
nt_up	numeric(1) indicating the number of nucleotides upstream the TSS to define the promoter region (1000 by default)
nt_down	numeric(1) indicating the number of nucleotides downstream the TSS to define the promoter region (200 by default)
include_normal_tissues	logical(1). If TRUE, the function will include normal peritumoral tissues in addition to tumoral samples. Default is FALSE.

Value

a Dataframe giving for each TCGA sample, the methylation level of a gene (mean methylation of probes located in its promoter) and the expression level of the gene (TPM value). The number of probes used to estimate the methylation level is also reported.

Examples

```
## Not run:
  CTexploreR:::prepare_TCGA_methylation_expression("LUAD", gene = "TDRD1")

## End(Not run)
```

---

set_fontsize *Determine font size*

Description

Gives the fontsize to use for the heatmap based on the matrix's dimension.

Usage

```
set_fontsize(matrix)
```

Arguments

matrix	matrix containing the data to visualise

Value

A logical number that is the fontsize to use

Examples

```
CTexploreR:::set_fontsize(matrix(1:3, 9,8))
```

subset_database *Subset databases*

Description

Check the presence of the genes in the database then subsets the database to only keep these genes' data.

Usage

```
subset_database(variable = NULL, data)
```

Arguments

variable	character() containing the names genes to keep in the data
data	Summarized Experiment or SingleCellExperiment object with valid variable names.

Value

A Summarized Experiment or SingleCellExperiment object with only the variables data

Examples

```
CTexploreR:::subset_database(variable = "MAGEA1", data = CTdata::GTEX_data())
```

TCGA_expression *Gene expression in TCGA tumors*

Description

Plots a heatmap of genes expression in TCGA samples (peritumoral and tumor samples when a specific tumor type is specified, or tumor samples only when tumor option is set to "all")

Usage

```
TCGA_expression(
  tumor = "all",
  genes = NULL,
  units = c("TPM", "log_TPM"),
  values_only = FALSE
)
```

Arguments

tumor	character defining the TCGA tumor type. Can be one of "SKCM", "LUAD", "LUSC", "COAD", "ESCA", "BRCA", "HNSC", or "all" (default).
genes	character nameing the selected genes. The default value, NULL, takes all CT genes.
units	character(1) with expression values unit. Can be "TPM" (default) or "log_TPM" (log(TPM + 1)).
values_only	logical(1). If TRUE, the function will return the expression values in all samples instead of the heatmap. Default is FALSE.

Value

A heatmap of selected CT genes expression in TCGA samples. If values_only = TRUE, TPM expression data is returned instead.

Examples

```
## Not run:
TCGA_expression(
    tumor = "LUAD", genes = c("MAGEA1", "MAGEA3"),
    units = "log_TPM")

## End(Not run)
```

TCGA_methylation_expression_correlation

Methylation-Expression correlation of Cancer-Testis genes in TCGA samples

---

Description

Plots the correlation between methylation and expression values of a Cancer-Testis (CT) gene in TCGA samples.

Usage

```
TCGA_methylation_expression_correlation(
  tumor = "all",
  gene = NULL,
  nt_up = 1000,
  nt_down = 200,
  min_probe_number = 3,
  include_normal_tissues = FALSE,
  values_only = FALSE
)
```

Arguments

tumor	character defining the TCGA tumor type. Can be one of "SKCM", "LUAD", "LUSC", "COAD", "ESCA", "BRCA", "HNSC", or "all" (default).
gene	character selected gene.
nt_up	numeric(1) indicating the number of nucleotides upstream the TSS to define the promoter region (1000 by default)
nt_down	numeric(1) indicating the number of nucleotides downstream the TSS to define the promoter region (200 by default)
min_probe_number	numeric(1) indicating the minimum number of probes (with methylation values) within the selected region to calculate its mean methylation level. Default is 3.
include_normal_tissues	logical(1). If TRUE, the function will include normal peritumoral tissues in addition to tumoral samples. Default is FALSE.
values_only	logical(1). If TRUE, the function will return the methylation and expression values in TCGA samples instead of the heatmap. Default is FALSE.

Details

The coefficient of correlation is set to NA if no probes are found in promoter regions or if less than 1% of tumors are positive (TPM >= 1) for the gene.

Value

A scatter plot representing for each TCGA sample, gene expression and mean methylation values of probe(s) located in its promoter region (defined as 1000 nucleotides upstream TSS and 200 nucleotides downstream TSS by default). If values_only = TRUE, methylation and expression values are returned in a tibble instead.

Examples

```
## Not run:
TCGA_methylation_expression_correlation("LUAD", gene = "TDRD1")

## End(Not run)
```

---

testis_expression *Gene expression in testis cells*

Description

Plots a heatmap of genes expression in the different types of testis cells, using scRNAseq data from "The adult human testis transcriptional cell atlas" (Guo et al. 2018)

Usage

```
testis_expression(
  cells = c("all", "germ_cells", "somatic_cells", "SSC", "Spermatogonia",
   "Early_spermatocyte", "Late_spermatocyte", "Round_spermatid", "Elongated_spermatid",
   "Sperm1", "Sperm2", "Macrophage", "Endothelial", "Myoid", "Sertoli", "Leydig"),
  genes = NULL,
  scale_lims = NULL,
  values_only = FALSE
)
```

Arguments

cells	character defining the testis cell types to be plotted. Can be "germ_cells", "somatic_cells", "all" (default), or any or a combination of "SSC", "Spermatogonia", "Early_spermatocyte", "Late_spermatocyte", "Round_spermatid", "Elongated_spermatid", "Sperm1", "Sperm2", "Macrophage", "Endothelial", "Myoid", "Sertoli", "Leydig".
genes	character nameing the selected genes. The default value, NULL, takes all CT genes.
scale_lims	vector of length 2 setting the lower and upper limits of the heatmap colorbar. By default, the lower limit is 0, and the upper limit corresponds to the third quartile of the logcounts values.
values_only	logical(1). If TRUE, the function will return the SingleCellExperiment instead of the heatmap. Default is FALSE.

Value

A heatmap of selected CT genes expression in single cells from adult testis. If values_only = TRUE, a SingleCellExperiment instead of the heatmap is returned instead.

Examples

```
## Not run:
testis_expression(cells = "germ_cells",
                  genes = c("MAGEA1", "MAGEA3", "MAGEA4"))

## End(Not run)
```

Index