Package 'TENET'

November 12, 2026

Title R package for TENET (Tracing regulatory Element Networks using Epigenetic Traits) to identify key transcription factors

Description
TENET identifies key transcription factors (TFs) and regulatory elements (REs) linked to a specific cell type by finding significantly correlated differences in gene expression and RE DNA methylation between case and control input datasets, and identifying the top genes by number of significant RE DNA methylation site links. It also includes many tools for visualization and analysis of the results, including plots displaying and comparing methylation and expression data and methylation site link counts, survival analysis, TF motif searching in the vicinity of linked RE DNA methylation sites, custom TAD and peak overlap analysis, and UCSC Genome Browser track file generation. A utility function is also provided to download methylation, expression, and patient survival data from The Cancer Genome Atlas (TCGA) for use in TENET or other analyses.

Version 1.4.0

URL https://github.com/rhielab/TENET

BugReports https://github.com/rhielab/TENET/issues

License GPL-2

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

BiocType Software

biocViews Software, BiomedicalInformatics, CellBiology, Genetics, Epigenetics, MultipleComparison, GeneExpression, DifferentialExpression, DNAMethylation, DifferentialMethylation, MethylationArray, Sequencing, MethylSeq, RNASeq, FunctionalGenomics, GeneRegulation, GeneTarget, HistoneModification, Transcription, Transcriptomics, Survival, Visualization

Depends R (>= 4.5)

Imports graphics, grDevices, stats, utils, tools, S4Vectors, GenomicRanges, IRanges, parallel, pastecs, ggplot2 (>= 4.0), RCircos, survival, BSgenome.Hsapiens.UCSC.hg38, seqLogo,

Biostrings, matlab, TCGAbiolinks, methods, R.utils, MultiAssayExperiment, SummarizedExperiment, sesame, sesameData, AnnotationHub, ExperimentHub, TENET.ExperimentHub, rtracklayer, MotifDb, BAMMtools, survminer

Suggests TENET.AnnotationHub, knitr, rmarkdown, BiocStyle

VignetteBuilder knitr

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

git_branch RELEASE_3_23

git_last_commit 0377fbe

git_last_commit_date 2026-04-28

Repository Bioconductor 3.23

Date/Publication 2026-11-11

Author Rhie Lab at the University of Southern California [cre], Daniel Mullen [aut] (ORCID: <https://orcid.org/0000-0002-7639-0549>), Zexun Wu [aut] (ORCID: <https://orcid.org/0000-0003-2566-1326>), Ethan Nelson-Moore [aut] (ORCID: <https://orcid.org/0009-0001-6903-9232>), Suhn Rhie [aut] (ORCID: <https://orcid.org/0000-0002-5522-5296>)

Maintainer Rhie Lab at the University of Southern California <rhielab@gmail.com>

Contents

easyTENET . . . 3
humanTranscriptionFactorDb . . . 6
humanTranscriptionFactorList . . . 8
step1MakeExternalDatasets . . . 8
step2GetDifferentiallyMethylatedSites . . . 11
step3GetAnalysisZScores . . . 15
step4SelectMostSignificantLinksPerDNAMethylationSite . . . 17
step5OptimizeLinks . . . 19
step6DNAMethylationSitesPerGeneTabulation . . . 21
step7ExpressionVsDNAMethylationScatterplots . . . 23
step7LinkedDNAMethylationSiteCountHistograms . . . 27
step7LinkedDNAMethylationSitesMotifSearching . . . 29
step7SelectedDNAMethylationSitesCaseVsControlBoxplots . . . 33
step7StatesForLinks . . . 35
step7TopGenesCaseVsControlExpressionBoxplots . . . 36
step7TopGenesCircosPlots . . . 39
step7TopGenesDNAMethylationHeatmaps . . . 41
step7TopGenesExpressionCorrelationHeatmaps . . . 43
step7TopGenesOverlappingLinkedDNAMethylationSitesHeatmaps . . . 45
step7TopGenesSurvival . . . 47
step7TopGenesTADTables . . . 52
step7TopGenesUCSCBedFiles . . . 55
step7TopGenesUserPeakOverlap . . . 58

TCGADownloader 61
TENET 64
TENETCacheAllData 65

Index **66**

`easyTENET` *Run the step 1 through step 6 functions with default arguments*

Description

This function runs the main six TENET functions (`step1MakeExternalDatasets`, `step2GetDifferentiallyMethylatedS`
`step3GetAnalysisZScores`, `step4SelectMostSignificantLinksPerDNAMethylationSite`, `step5OptimizeLinks`,
and `step6DNAMethylationSitesPerGeneTabulation`) in sequence on the specified TENETMultiAssayExperiment object. Arguments for this function generally reflect the arguments of the component functions without clearly defined defaults, with the exception of the `step1MakeExternalDatasets` function where all arguments have been included to support all available options to define regions with relevant regulatory elements. All remaining arguments of the component functions are set to their default values.

Usage

```
easyTENET(
  TENETMultiAssayExperiment,
  extHM = NA,
  extNDR = NA,
  consensusEnhancer = TRUE,
  consensusPromoter = FALSE,
  consensusNDR = TRUE,
  publicEnhancer = FALSE,
  publicPromoter = FALSE,
  publicNDR = FALSE,
  cancerType = NA,
  ENCODEPLS = FALSE,
  ENCODEpELS = FALSE,
  ENCODEdELS = FALSE,
  assessPromoter = FALSE,
  TSSDist = 1500,
  minCaseCount,
  coreCount = 1
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. Coordinates for genes and DNA methylation sites must be included in

the rowRanges of their respective SummarizedExperiment objects and should be annotated to the human hg38 genome. An argument of all functions except `step1MakeExternalDatasets`.

`extHM` To use custom histone modification datasets, specify one or more paths to .bed, .narrowPeak, .broadPeak, and/or .gappedPeak files containing these datasets, or directories containing these file types. The files may optionally be compressed (.gz/.bz2/.xz). Otherwise, specify NA or do not specify this argument. An argument of the `step1MakeExternalDatasets` function.

`extNDR` To use custom open chromatin or NDR datasets, specify one or more paths to .bed, .narrowPeak, .broadPeak, and/or .gappedPeak files containing these datasets, or directories containing these file types. The files may optionally be compressed (.gz/.bz2/.xz). Otherwise, specify NA or do not specify this argument. An argument of the `step1MakeExternalDatasets` function.

`consensusEnhancer`
Set to TRUE to use the consensus enhancer data included in TENET.AnnotationHub. Defaults to TRUE. An argument of the `step1MakeExternalDatasets` function.

`consensusPromoter`
Set to TRUE to use the consensus promoter data included in TENET.AnnotationHub. Defaults to FALSE. An argument of the `step1MakeExternalDatasets` function.

`consensusNDR` Set to TRUE to use the consensus open chromatin (NDR) data included in TENET.AnnotationHub. Defaults to TRUE. An argument of the `step1MakeExternalDatasets` function.

`publicEnhancer` Set to TRUE to use the preprocessed publicly available enhancer (H3K27ac) datasets included in TENET.AnnotationHub. If set to TRUE, `cancerType` must be specified. Defaults to FALSE. An argument of the `step1MakeExternalDatasets` function.

`publicPromoter` Set to TRUE to use the preprocessed publicly available promoter (H3K4me3) datasets included in TENET.AnnotationHub. If set to TRUE, `cancerType` must be specified. Defaults to FALSE. An argument of the `step1MakeExternalDatasets` function.

`publicNDR` Set to TRUE to use the preprocessed publicly available open chromatin (ATAC-seq, DNase-seq) datasets included in TENET.AnnotationHub. If set to TRUE, `cancerType` must be specified. Defaults to FALSE. An argument of the `step1MakeExternalDatasets` function.

`cancerType` If `publicEnhancer`, `publicPromoter`, and/or `publicNDR` is TRUE, specify a vector of cancer types ('BLCA', 'BRCA', 'COAD', 'ESCA', 'HNSC', 'KIRP', 'LIHC', 'LUAD', 'LUSC', and/or 'THCA') to include the public data relevant to those cancer types. Defaults to NA. An argument of the `step1MakeExternalDatasets` function.

`ENCODEPLS` Set to TRUE to use the ENCODE promoter-like elements dataset included in TENET.AnnotationHub. Defaults to FALSE. An argument of the `step1MakeExternalDatasets` function.

`ENCODEpELS` Set to TRUE to use the ENCODE proximal enhancer-like elements dataset included in TENET.AnnotationHub. Defaults to FALSE. An argument of the `step1MakeExternalDatasets` function.

ENCODEdELS — Set to TRUE to use the ENCODE distal enhancer-like elements dataset included in TENET.AnnotationHub. Defaults to FALSE. An argument of the step1MakeExternalDatasets function.

assessPromoter — Set to TRUE to identify DNA methylation sites that mark promoter regions or FALSE to identify distal enhancer regions. Defaults to FALSE. An argument of the step2GetDifferentiallyMethylatedSites function.

TSSDist — Specify a positive integer distance in base pairs to any transcription start site within which DNA methylation sites are considered promoter DNA methylation sites. DNA methylation sites outside this distance from any transcription start site will be considered enhancer methylation sites. Defaults to 1500. An argument of the step2GetDifferentiallyMethylatedSites function.

minCaseCount — Specify the minimum number of case samples to be considered for the hyper- and hypomethylated groups. Must be a positive integer less than the total number of case samples. An argument of the step2GetDifferentiallyMethylatedSites function.

coreCount — Argument passed as the mc.cores argument to mclapply. See ?parallel::mclapply for more details. Defaults to 1. Used by the step3GetAnalysisZScores, step4SelectMostSignificantLinksPerDNAMethylationSite, and step5OptimizeLinks functions.

Value

Returns the created MultiAssayExperiment object containing data from all step 1 through 6 functions.

Examples

```
## This example creates a dataset of putative enhancer regulatory elements
## from consensus datasets and breast invasive carcinoma-relevant sources
## collected in the TENET.AnnotationHub package, then runs the step 2 through
## step 6 TENET functions analyzing RE DNA methylation sites in potential
## enhancer elements located over 1500 bp from transcription start sites
## listed for genes and transcripts in the GENCODE v36 human genome
## annotations, using a minimum case sample count of 5 and one CPU core
## to perform the analysis.

## Load the example TENET MultiAssayExperiment object from the
## TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example TENET MultiAssayExperiment to run the step 1 through
## step 6 TENET functions
returnValue <- easyTENET(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    extHM = NA,
    extNDR = NA,
    publicEnhancer = TRUE,
    publicNDR = TRUE,
    cancerType = "BRCA",
```

```
    ENCODEdELS = TRUE,
    minCaseCount = 5
)

## This example creates a dataset of putative promoter regulatory elements
## using BED-like files contained in the user's working directory, consensus
## NDR and promoter regions, and regions with promoter-like signatures from
## the ENCODE SCREEN project, but excluding cancer type-specific public
## datasets. This dataset is then used to analyze DNA methylation sites in
## promoter elements within 2000 bp of all transcription start sites
## provided in the MultiAssayExperiment only, identifying alterations found
## in at least 10 samples, and using 8 CPU cores to perform the analysis.

## Load the example TENET MultiAssayExperiment object from the
## TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example TENET MultiAssayExperiment to run the step 1 through
## step 6 TENET functions
returnValue <- easyTENET(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    extHM = ".",
    extNDR = ".",
    consensusEnhancer = FALSE,
    consensusPromoter = TRUE,
    ENCODEPLS = TRUE,
    assessPromoter = TRUE,
    TSSDist = 2000,
    minCaseCount = 10,
    coreCount = 8
)
```

`humanTranscriptionFactorDb`

Human transcription factor database

Description

A data frame with information on genes identified as human TFs by Lambert SA et al (PMID: 29425488). Candidate proteins were manually examined by a panel of experts based on available data. Proteins with experimentally demonstrated DNA binding specificity were considered TFs. Other proteins, such as co-factors and RNA binding proteins, were classified as non-TFs. **Citation:** Lambert SA, Jolma A, Campitelli LF, et al. The Human Transcription Factors. Cell. 2018 Feb 8;172(4):650-665. doi: 10.1016/j.cell.2018.01.029. Erratum in: Cell. 2018 Oct 4;175(2):598-599. PMID: 29425488.

Usage

```
data("humanTranscriptionFactorDb", package = "TENET")
```

Format

A data frame with 2765 rows and 28 variables.

`Ensembl.ID` character Official Ensembl gene ID.

`HGNC.symbol` character Official gene name.

`DBD` character DNA binding domains contained in protein(s).

`Is.TF.` character Is the protein a TF (Yes/No).

`TF.assessment` character Assessment of binding activity.

`Binding.mode` character Mode of interacting with DNA.

`Motif.status` character Current status of motif availability.

`Final.Notes` character Final notes, automatically generated.

`Final.Comments` character Final comments, manually entered.

`Interpro.ID.s.` character Interpro IDs for DBDs.

`EntrezGene.ID` character Entrez Gene ID.

`EntrezGene.Description` character Entrez Gene Description.

`PDB.ID` character Protein Data Bank ID.

`TF.tested.by.HT.SELEX.` character Has the protein been tested for DNA binding in a HT-SELEX assay in the Taipale lab?

`TF.tested.by.PBM.` character Has the protein been tested for DNA binding in a PBM assay?

`Conditional.Binding.Requirements` character Notes on requirements for binding.

`Original.Comments` character Original comments provided by the primary reviewer of the protein.

`Vaquerizas.2009.classification` character Classification provided by the Vaquerizas 2009 paper.

`CisBP.considers.it.a.TF.` character Is the protein available in the CisBP database (build 1.02)?

`TFCat.classification` character Does the TFCat web site classify the protein as a TF?

`Is.a.GO.TF.` character Does GO (Gene Ontology) classify the protein as a TF?

`Initial.assessment` character Initial assessment provided by curators.

`Curator.1` character Name of curator 1.

`Curator.2` character Name of curator 2.

`TFclass.considers.it.a.TF.` character Does TFclass consider the protein to be a TF?

`Go.Evidence` character Evidence from GO supporting this protein being a TF.

`Pfam.Domains..By.ENSP.ID.` character List of Pfam Domains contained in the protein.

`Is.C2H2.ZF.KRAB..` logical Is the protein a KRAB-containing Cys2-His2 zinc finger (C2H2-ZF) protein? **Note:** This description is a guess; Lambert et al did not provide a description for this field.

Source

http://humantfs.ccbr.utoronto.ca/download.php

Examples

```
data("humanTranscriptionFactorDb", package = "TENET")
```

humanTranscriptionFactorList *Human transcription factor list*

Description

A character vector of the Ensembl IDs of genes identified as human TFs by Lambert SA et al (PMID: 29425488). Candidate proteins were manually examined by a panel of experts based on available data. Proteins with experimentally demonstrated DNA binding specificity were considered TFs. Other proteins, such as co-factors and RNA binding proteins, were classified as non-TFs. **Citation:** Lambert SA, Jolma A, Campitelli LF, et al. The Human Transcription Factors. Cell. 2018 Feb 8;172(4):650-665. doi: 10.1016/j.cell.2018.01.029. Erratum in: Cell. 2018 Oct 4;175(2):598-599. PMID: 29425488.

Usage

```
data("humanTranscriptionFactorList", package = "TENET")
```

Format

A character vector containing 1,639 Ensembl IDs of known human TFs.

Source

http://humantfs.ccbr.utoronto.ca/download.php

Examples

```
data("humanTranscriptionFactorList", package = "TENET")
```

step1MakeExternalDatasets *Create a GRanges object representing putative regulatory element regions, based on the data sources selected for inclusion, to be used in later TENET steps*

Description

This function creates a GRanges object containing regions representing putative regulatory elements, either enhancers or promoters, of interest to the user, based on the presence of specific histone marks and open chromatin/nucleosome-depleted regions. This function can take input from user-specified BED-like files (see https://genome.ucsc.edu/FAQ/FAQformat.html#format1) containing regions with histone modification (via the `extHM` argument) and/or open chromatin/nucleosome-depleted regions (via the `extNDR` argument), as well as preprocessed enhancer, promoter, and open chromatin datasets from many cell/tissue types included in the TENET.AnnotationHub repository.

The resulting GRanges object will be returned. GRanges objects created by this function can be used by the `step2GetDifferentiallyMethylatedSites` function or other downstream functions. **Note:** Using datasets from TENET.AnnotationHub requires an internet connection, as those datasets are hosted in the Bioconductor AnnotationHub Data Lake.

Usage

```
step1MakeExternalDatasets(
  extHM = NA,
  extNDR = NA,
  consensusEnhancer = TRUE,
  consensusPromoter = FALSE,
  consensusNDR = TRUE,
  publicEnhancer = FALSE,
  publicPromoter = FALSE,
  publicNDR = FALSE,
  cancerType = NA,
  ENCODEPLS = FALSE,
  ENCODEpELS = FALSE,
  ENCODEdELS = FALSE
)
```

Arguments

`extHM`	To use custom histone modification datasets, specify one or more paths to .bed, .narrowPeak, .broadPeak, and/or .gappedPeak files containing these datasets, or directories containing these file types. The files may optionally be compressed (.gz/.bz2/.xz). Otherwise, specify NA or do not specify this argument.
`extNDR`	To use custom open chromatin or NDR datasets, specify one or more paths to .bed, .narrowPeak, .broadPeak, and/or .gappedPeak files containing these datasets, or directories containing these file types. The files may optionally be compressed (.gz/.bz2/.xz). Otherwise, specify NA or do not specify this argument.
`consensusEnhancer`	Set to TRUE to use the consensus enhancer data included in TENET.AnnotationHub. Defaults to TRUE.
`consensusPromoter`	Set to TRUE to use the consensus promoter data included in TENET.AnnotationHub. Defaults to FALSE.
`consensusNDR`	Set to TRUE to use the consensus open chromatin (NDR) data included in TENET.AnnotationHub. Defaults to TRUE.
`publicEnhancer`	Set to TRUE to use the preprocessed publicly available enhancer (H3K27ac) datasets included in TENET.AnnotationHub. If set to TRUE, `cancerType` must be specified. Defaults to FALSE.
`publicPromoter`	Set to TRUE to use the preprocessed publicly available promoter (H3K4me3) datasets included in TENET.AnnotationHub. If set to TRUE, `cancerType` must be specified. Defaults to FALSE.

publicNDR — Set to TRUE to use the preprocessed publicly available open chromatin (ATAC-seq, DNase-seq) datasets included in TENET.AnnotationHub. If set to TRUE, cancerType must be specified. Defaults to FALSE.

cancerType — If publicEnhancer, publicPromoter, and/or publicNDR is TRUE, specify a vector of cancer types ('BLCA', 'BRCA', 'COAD', 'ESCA', 'HNSC', 'KIRP', 'LIHC', 'LUAD', 'LUSC', and/or 'THCA') to include the public data relevant to those cancer types. Defaults to NA.

ENCODEPLS — Set to TRUE to use the ENCODE promoter-like elements dataset included in TENET.AnnotationHub. Defaults to FALSE.

ENCODEpELS — Set to TRUE to use the ENCODE proximal enhancer-like elements dataset included in TENET.AnnotationHub. Defaults to FALSE.

ENCODEdELS — Set to TRUE to use the ENCODE distal enhancer-like elements dataset included in TENET.AnnotationHub. Defaults to FALSE.

Value

Returns the created regulatory element GRanges object.

Examples

```
## This example creates a dataset of putative enhancer regulatory elements
## from consensus datasets and breast invasive carcinoma-relevant sources
## collected in the TENET.AnnotationHub package.
returnGRanges <- step1MakeExternalDatasets(
    extHM = NA,
    extNDR = NA,
    publicEnhancer = TRUE,
    publicNDR = TRUE,
    cancerType = "BRCA",
    ENCODEdELS = TRUE
)

## This example creates a dataset of putative promoter regulatory elements
## using user provided BED-like files contained in the working
## directory, consensus NDR and promoter regions, and regions with
## promoter-like signatures from the ENCODE SCREEN project. This excludes any
## cancer type-specific public datasets.
returnGRanges <- step1MakeExternalDatasets(
    extHM = ".",
    extNDR = ".",
    consensusEnhancer = FALSE,
    consensusPromoter = TRUE,
    ENCODEPLS = TRUE
)
```

`step2GetDifferentiallyMethylatedSites` *Identify differentially methylated RE DNA methylation sites*

Description

This function identifies DNA methylation sites that mark putative regulatory elements (REs), including enhancer and promoter regions. These are sites that lie within regions from a user-supplied GRanges object, such as one created by the `step1MakeExternalDatasets` function, and which are located at a user-specified distance relative to the transcription start sites (TSS) listed in either the rowRanges of the elementMetadata of the "expression" SummarizedExperiment in the TENETMultiAssayExperiment object, or the selected `geneAnnotationDataset` (which will be filtered to only genes and transcripts). After identifying DNA methylation sites representing the specified REs, the function classifies the RE DNA methylation sites as methylated, unmethylated, hypermethylated, or hypomethylated based on their differential methylation between the control and case samples supplied by the user, defined by cutoff values which are either automatically based on the mean methylation densities of the identified RE DNA methylation sites, or manually set by the user. **Note:** Using the algorithm to set cutoffs is recommended for use with DNA methylation array data, and may not work for whole-genome DNA methylation data.

Usage

```
step2GetDifferentiallyMethylatedSites(
  TENETMultiAssayExperiment,
  regulatoryElementGRanges = NA,
  geneAnnotationDataset = NA,
  DNAMethylationArray = NA,
  assessPromoter = FALSE,
  TSSDist = 1500,
  purityData = NA,
  methCutoff = NA,
  hypomethCutoff = NA,
  hypermethCutoff = NA,
  unmethCutoff = NA,
  methUnmethProportionOffset = 0.2,
  hypomethHypermethProportionOffset = 0.1,
  minCaseCount,
  cgDNAMethylationSitesOnly = TRUE
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. Coordinates for genes and DNA methylation sites must be included in the rowRanges of their respective SummarizedExperiment objects and should be

annotated to the same genome build as the regions given in the regulatoryElementGRanges object.

`regulatoryElementGRanges`
: Specify a GRanges object containing genomic regions representing regulatory elements of interest to the user. Coordinates for the regulatory element regions should be annotated to the same genome build as the gene and DNA methylation site coordinates given in the TENETMultiAssayExperiment object. If this argument is set to NA or not specified, this function will use all DNA methylation sites representing regulatory elements of interest as defined by the assessPromoter and TSSDist arguments. Defaults to NA.

`geneAnnotationDataset`
: Specify a gene annotation dataset which is used to identify transcription start sites in order to find DNA methylation sites within regulatory elements of interest (promoters or enhancers) in conjunction with the settings of the assessPromoter and TSSDist arguments. The dataset will be filtered to only genes and transcripts. The argument must be either a GRanges object (such as one imported via `rtracklayer::import`) or a path to a GFF3 or GTF file. Both GENCODE and Ensembl annotations are supported. Other annotation datasets may work, but have not been tested. See the "Input data" section of the vignette for information on the required dataset format. Specify NA to use the start coordinates of all entries in the elementMetadata of the rowRanges of the "expression" SummarizedExperiment object within the TENETMultiAssayExperiment object, in which case no filtering will be done and all entries will be assumed to represent transcripts. Defaults to NA.

`DNAMethylationArray`
: Specify the name of a DNA methylation probe array supported by the sesameData package (see `?sesameData::sesameData_getManifestGRanges`). If an array is specified, RE DNA methylation sites and their locations in that array's manifest are cross-referenced with RE DNA methylation site IDs included in the rownames of the methylation dataset provided in the "methylation" SummarizedExperiment object within the TENETMultiAssayExperiment object, and only those overlapping will be considered for analysis. If set to NA, all RE DNA methylation sites with locations listed in the rowRanges of the "methylation" SummarizedExperiment object are used. Defaults to NA.

`assessPromoter`
: Set to TRUE to identify DNA methylation sites that mark promoter regions or FALSE to identify distal enhancer regions. Defaults to FALSE.

`TSSDist`
: Specify a positive integer distance in base pairs to any transcription start site (see `geneAnnotationDataset`) within which DNA methylation sites are considered promoter DNA methylation sites. DNA methylation sites outside this distance from any transcription start site will be considered enhancer methylation sites. Defaults to 1500.

`purityData`
: Specify a SummarizedExperiment object which contains DNA methylation datasets collected from potential cell types which might affect the purity of the patient samples contained in the TENETMultiAssayExperiment. The coordinates for DNA methylation sites in this dataset should be included in the rowRanges of the purityData SummarizedExperiment object. Additionally, the DNA methylation site IDs in the purityData SummarizedExperiment object should overlap

with DNA methylation sites present in the TENETMultiAssayExperiment and only those that do overlap will be considered for analysis. Defaults to NA.

`methCutoff` Specify a number from 0 to 1 to be the beta-value cutoff for methylated RE DNA methylation sites. If unspecified or NA, an algorithm will be used to find the optimal cutoff value.

`hypomethCutoff` Specify a number from 0 to 1 to be the beta-value cutoff for hypomethylated RE DNA methylation sites. Should be set lower than the methCutoff. If unspecified or NA, an algorithm will be used to find the optimal cutoff value.

`hypermethCutoff`
Specify a number from 0 to 1 to be the beta-value cutoff for hypermethylated RE DNA methylation sites. Should be set higher than the unmethCutoff. If unspecified or NA, an algorithm will be used to find the optimal cutoff value.

`unmethCutoff` Specify a number from 0 to 1 to be the beta-value cutoff for unmethylated RE DNA methylation sites. If unspecified or NA, an algorithm will be used to find the optimal cutoff value.

`methUnmethProportionOffset`
Specify a number from 0 to 1 indicating a proportion of the size of the region between the first and last local maxima in the density plot of the mean methylation values of the RE DNA methylation sites in the control samples. This proportion will be added to or subtracted from the position of these local maxima to set the unmethylation and methylation cutoffs, respectively, if they are not defined by the user. Ideally should not exceed 0.5. Defaults to 0.2.

`hypomethHypermethProportionOffset`
Specify a number from 0 to 1 indicating a proportion of the size of the region between the first and last local maxima in the density plot of the mean methylation values of the RE DNA methylation sites in the case samples. This proportion will be added to or subtracted from the calculated unmethylation and methylation cutoffs to set the hypermethylation and hypomethylation cutoffs, respectively, if they are not defined by the user. Ideally should not exceed 0.5. Defaults to 0.1.

`minCaseCount` Specify the minimum number of case samples to be considered for the hyper- and/or hypomethylated groups. Must be a positive integer less than the total number of case samples.

`cgDNAMethylationSitesOnly`
Set to TRUE to include only RE DNA methylation sites with IDs that start with "cg". TRUE means that RE DNA methylation sites whose IDs do not start with "cg" will be removed from TENET analyses. Defaults to TRUE.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named "step2GetDifferentiallyMethylatedSites" in its metadata containing the output of this function. These data include the set of calculated cutoff values, the identities and counts of the classified RE DNA methylation sites, as well as plots of the mean methylation distributions of the identified regulatory element DNA methylation sites in the case and control samples and the set cutoff values. **Note:** If assessPromoter is TRUE, two distribution plots are saved, one using all promoter DNA methylation sites, and one using only promoter DNA methylation sites which are identified to overlap REs.

Examples

```
## This example uses datasets provided in the TENET.ExperimentHub package to
## perform an example analysis, considering RE DNA methylation sites in
## potential enhancer elements located over 1500 bp from transcription
## start sites listed for genes and transcripts in the GENCODE v36 human
## genome annotations, using a minimum case sample count of 5, and otherwise
## using default settings.

## Load the example TENET MultiAssayExperiment object, and the example
## GRanges object created by the TENET step 1 function, from the
## TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()
exampleStep1MakeExternalDatasetsGRangesObject <-
    TENET.ExperimentHub::exampleTENETStep1MakeExternalDatasetsGRanges()

## Use the example datasets to identify differentially methylated
## RE DNA methylation sites
returnValue <- step2GetDifferentiallyMethylatedSites(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    regulatoryElementGRanges =
        exampleTENETStep1MakeExternalDatasetsGRanges,
    minCaseCount = 5
)

## This example uses the same datasets, this time analyzing DNA methylation
## sites in promoter elements, considering all RE DNA methylation sites
## found within 2000 bp of only the transcription start sites provided in the
## MultiAssayExperiment. All methylation cutoffs are manually specified, the
## minimum case sample count is set to 10, and all RE DNA methylation sites
## are considered regardless of whether their IDs begin with "cg".

## Load the example TENET MultiAssayExperiment object, and the example
## GRanges object created by the TENET step 1 function, from the
## TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()
exampleStep1MakeExternalDatasetsGRangesObject <-
    TENET.ExperimentHub::exampleTENETStep1MakeExternalDatasetsGRanges()

## Use the example datasets to identify differentially methylated
## RE DNA methylation sites
returnValue <- step2GetDifferentiallyMethylatedSites(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    regulatoryElementGRanges =
        exampleTENETStep1MakeExternalDatasetsGRanges,
    geneAnnotationDataset = NA,
    assessPromoter = TRUE,
    TSSDist = 2000,
    methCutoff = 0.8,
    hypomethCutoff = 0.7,
    hypermethCutoff = 0.3,
```

```
    unmethCutoff = 0.2,
    minCaseCount = 10,
    cgDNAMethylationSitesOnly = FALSE
)
```

`step3GetAnalysisZScores` *Calculate Z-scores comparing the mean expression of each gene in the case samples that are hyper- and/or hypomethylated for each RE DNA methylation site identified in step 2*

Description

This function calculates Z-scores comparing the mean expression of each gene in the case samples that are hyper- and/or hypomethylated for each RE DNA methylation site identified in step 2, according to the methylation cutoffs set in step 2, to the mean expression of the remaining non-hyper- or hypomethylated case samples. By identifying significant Z-scores, initial RE DNA methylation site-gene links are identified, in the form of case samples with hyper- or hypomethylation of a particular RE DNA methylation site also displaying particularly high or low expression of specific genes.

Usage

```
step3GetAnalysisZScores(
  TENETMultiAssayExperiment,
  hypermethAnalysis = TRUE,
  hypomethAnalysis = TRUE,
  includeControl = FALSE,
  TFOnly = TRUE,
  zScoreCalculation = "oneSample",
  sparseResults = TRUE,
  pValue = 0.05,
  coreCount = 1
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step2GetDifferentiallyMethylated` function.

`hypermethAnalysis`
: Set to TRUE to calculate Z-scores for hypermethylated RE DNA methylation sites. Defaults to TRUE.

`hypomethAnalysis`
: Set to TRUE to calculate Z-scores for hypomethylated RE DNA methylation sites. Defaults to TRUE.

`includeControl`
: Set to TRUE to include the control samples when identifying hyper/hypomethylated groups and calculating Z-scores. Defaults to FALSE.

`TFOnly`
: Set to TRUE to only consider genes that are accepted transcription factors according to "The Human Transcription Factors" by Lambert et al. 2018 when calculating Z-scores. Defaults to TRUE.

`zScoreCalculation`
: Set to 'oneSample' to use a one-sample Z-score calculation or 'twoSample' to use a two sample Z-score calculation. Note that 'twoSample' tends to be much more lenient, and identifies many more significant RE DNA methylation site-gene links. Defaults to 'oneSample'.

`sparseResults`
: Set to TRUE to save only the significant Z-scores for RE DNA methylation site-gene links. **Note:** If multiple testing correction will be performed in the subsequent `step4SelectMostSignificantLinksPerDNAMethylationSite` function, this argument should be set to FALSE. Defaults to TRUE.

`pValue`
: Specify the p-value below which Z-scores will be considered significant during comparison of gene expression values between case samples that are hyper- or hypomethylated and those that are not. If `sparseResults` is set to TRUE, only significant Z-scores will be saved in the output MultiAssayExperiment object. Defaults to 0.05.

`coreCount`
: Argument passed as the mc.cores argument to mclapply. See `?parallel::mclapply` for more details. Defaults to 1.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named "step3GetAnalysisZScores" in its metadata containing the output of this function, which includes Z-scores comparing the mean expression of each gene in samples that are hypo- and/or hypomethylated for each RE DNA methylation site with the mean expression in samples that are not.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to calculate one-sample Z-scores for links
## between both hypermethylated and hypomethylated RE DNA methylation
## sites and the expression of transcription factor genes only, considering
## only case samples. Only significant Z-scores (based on a threshold of
## p<0.05) will be saved to the TENETMultiAssayExperiment object. The
## analysis will be performed using one CPU core.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Calculate Z-scores for hyper- and hypomethylated RE DNA methylation sites
```

```
returnValue <- step3GetAnalysisZScores(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment
)

## This example demonstrates many of the analysis options. It calculates
## two-sample Z-scores for links between only hypomethylated RE DNA
## methylation sites and all genes, considering both case and control samples
## All Z-scores will be saved to the TENETMultiAssayExperiment object
## (which takes a large amount of memory). Z-scores with p-values less
## than 0.1 will be considered significant. The analysis will be performed
## using 8 CPU cores.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Calculate Z-scores for only hypomethylated RE DNA methylation sites
returnValue <- step3GetAnalysisZScores(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    hypermethAnalysis = FALSE,
    includeControl = TRUE,
    TFOnly = FALSE,
    zScoreCalculation = "twoSample",
    sparseResults = FALSE,
    pValue = 0.1,
    coreCount = 8
)
```

`step4SelectMostSignificantLinksPerDNAMethylationSite` *Select the most significant RE DNA methylation site-gene links to each RE DNA methylation site*

Description

This function takes the calculated Z-scores for the hyper- and/or hypomethylated G+ RE DNA methylation site-gene links and selects the most significant links to each RE DNA methylation site, either up to a number specified by the user, or based on a significant p-value level set by the user after multiple testing correction is performed on the Z-scores output by the `step3GetAnalysisZScores` function per RE DNA methylation site in the RE DNA methylation site-gene pairs.

Usage

```
step4SelectMostSignificantLinksPerDNAMethylationSite(
  TENETMultiAssayExperiment,
  hypermethGplusAnalysis = TRUE,
  hypomethGplusAnalysis = TRUE,
```

```
  linksPerREDNAMethylationSiteMaximum = 25,
  multipleTestingCorrectionMethod = NA,
  multipleTestingPValue = 0.05,
  coreCount = 1
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step2GetDifferentiallyMethylated` and `step3GetAnalysisZScores` functions.

`hypermethGplusAnalysis`
: Set to TRUE to analyze hypermethylated G+ RE DNA methylation site-gene links. Requires the hypermethAnalysis parameter to have been set to TRUE in step 3.

`hypomethGplusAnalysis`
: Set to TRUE to analyze hypomethylated G+ RE DNA methylation site-gene links. Requires the hypomethAnalysis parameter to have been set to TRUE in step 3.

`linksPerREDNAMethylationSiteMaximum`
: This parameter must either be set to an integer n greater than 0, in which case only the n most significant RE DNA methylation site-gene link pairs from step 3 will be selected per unique RE DNA methylation site, or NA if using the multipleTestingPValue argument to set a significant p-value cutoff. Defaults to 25.

`multipleTestingCorrectionMethod`
: Specify a character string describing a multiple testing correction method supported by `p.adjust` (see `?stats::p.adjust`) to perform multiple testing correction on the Z-scores from step 3, using the `multipleTestingPValue` argument to specify the significant p-value cutoff, or specify NA to skip multiple testing correction, in which case `linksPerREDNAMethylationSiteMaximum` will be used to determine the number of links to retain. If specified, `linksPerREDNAMethylationSiteMaximu` will be ignored. Defaults to NA.

`multipleTestingPValue`
: Cutoff for multiple testing corrected p-values. This argument is only used if the `multipleTestingCorrectionMethod` argument is specified. Defaults to 0.05.

`coreCount`
: Argument passed as the mc.cores argument to mclapply. See `?parallel::mclapply` for more details. Defaults to 1.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named "step4SelectMostSignificantLinksPerDNAMethylationSite" in its metadata containing the most significant selected gene links to the hyper- and/or hypomethylated RE DNA methylation sites.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to identify the 25 most significant links
## between both hyper- and hypomethylated enhancer DNA methylation sites and
## all genes, using one CPU core to perform the analysis.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Perform the analysis
returnValue <- step4SelectMostSignificantLinksPerDNAMethylationSite(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment
)

## This example demonstrates many of the analysis options. It identifies
## the most significant links between only hypomethylated enhancer DNA
## methylation sites and all genes by performing Bonferroni multiple testing
## correction using a significant p-value of 0.10, using 8 CPU cores to
## perform the analysis. Note: Running this code with the
## exampleTENETMultiAssayExperiment will produce a warning message because
## sparseResults was set to TRUE when the example dataset was generated, but
## it is still valid as an example.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Perform the analysis
returnValue <- step4SelectMostSignificantLinksPerDNAMethylationSite(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    hypermethGplusAnalysis = FALSE,
    multipleTestingCorrectionMethod = "bonferroni",
    multipleTestingPValue = 0.1,
    coreCount = 8
)
```

---

step5OptimizeLinks *Find final RE DNA methylation site-gene links using various optimization metrics*

Description

This function takes the most significant hyper- and/or hypomethylated G+ RE DNA methylation site-gene links selected in step 4, and selects optimized links based on the relative expression of the given gene in hyper- or hypomethylated case samples compared to control samples, using an unpaired two-sided Wilcoxon rank-sum test to check that the hyper- or hypomethylated samples

for that given RE DNA methylation site-gene link also show appropriately higher/lower expression of the linked gene in a number of case samples greater than or equal to the `minCaseCount` number specified in the `step2GetDifferentiallyMethylatedSites` function that have maximum/minimum methylation above/below the `hyperStringency`/`hypoStringency` cutoff values selected.

Usage

```
step5OptimizeLinks(
  TENETMultiAssayExperiment,
  hypermethGplusAnalysis = TRUE,
  hypomethGplusAnalysis = TRUE,
  expressionPvalCutoff = 0.05,
  hyperStringency = NA,
  hypoStringency = NA,
  coreCount = 1
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step2GetDifferentiallyMethylated`
`step3GetAnalysisZScores`, and `step4SelectMostSignificantLinksPerDNAMethylationSite` functions.

`hypermethGplusAnalysis`
: Set to TRUE to optimize hypermethylated G+ links. Requires `hypermethAnalysis` to have been set to TRUE in step 4.

`hypomethGplusAnalysis`
: Set to TRUE to optimize hypomethylated G+ links. Requires `hypomethAnalysis` to have been set to TRUE in step 4.

`expressionPvalCutoff`
: Cutoff for Benjamini-Hochberg corrected Wilcoxon p-values used during comparison of gene expression values between hyper/hypomethylated case and control samples. Defaults to 0.05.

`hyperStringency`
: Specify a beta-value cutoff as a number from 0 to 1 to select hypermethylated links with methylation values above the cutoff. Defaults to the `hypermethCutoff` value specified in step 2.

`hypoStringency`
: Specify a beta-value cutoff as a number from 0 to 1 to select hypomethylated links with methylation values below the cutoff. Defaults to the `hypomethCutoff` value specified in step 2.

`coreCount`
: Argument passed as the mc.cores argument to mcmapply. See `?parallel::mcmapply` for more details. Defaults to 1.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named "step5OptimizeLinks" in its metadata containing the output of this function, which includes the gene and RE DNA methylation site IDs of each optimized gene-RE DNA methylation site link, the Z-scores and corresponding p-values calculated in steps 3-4 for those links, and the various optimization metrics calculated by this step.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to optimize links between the hypermethylated
## and hypomethylated RE DNA methylation sites and the expression of
## transcription factor genes. The default p-value cutoff of 0.05 is used,
## and the `hyperStringency` and `hypoStringency` values are set to the
## originally calculated hyper- and hypomethylation cutoffs. The analysis
## uses one CPU core.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Perform the link optimization
returnValue <- step5OptimizeLinks(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment
)

## This example demonstrates many of the analysis options. It only runs
## on hypomethylated RE DNA methylation sites and uses an expression p-value
## cutoff of 0.01. The `hypoStringency` value is set to 0.3, overriding the
## hypomethylation cutoff set in step 2. The analysis uses eight CPU cores.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Perform the link optimization
returnValue <- step5OptimizeLinks(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    hypermethGplusAnalysis = FALSE,
    expressionPvalCutoff = 0.01,
    hypoStringency = 0.3,
    coreCount = 8
)
```

`step6DNAMethylationSitesPerGeneTabulation` *Tabulate the total number of RE DNA methylation sites linked to each gene*

Description

This function takes the final optimized RE DNA methylation site-gene links identified in step 5 and tabulates the number of links per gene separately for the hyper- and/or hypomethylated G+ analysis quadrants.

Usage

```
step6DNAMethylationSitesPerGeneTabulation(
  TENETMultiAssayExperiment,
  geneAnnotationDataset = NA,
  hypermethGplusAnalysis = TRUE,
  hypomethGplusAnalysis = TRUE
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step5OptimizeLinks` function.

`geneAnnotationDataset`
: Specify a gene annotation dataset which is used to identify names for genes by their Ensembl IDs. The argument must be either a GRanges object (such as one imported via `rtracklayer::import`) or a path to a GFF3 or GTF file. Both GENCODE and Ensembl annotations are supported. Other annotation datasets may work, but have not been tested. See the "Input data" section of the vignette for information on the required dataset format. Specify NA to use the gene names listed in the "geneName" column of the elementMetadata of the rowRanges of the "expression" SummarizedExperiment object within the TENETMultiAssayExperiment object. Defaults to NA.

`hypermethGplusAnalysis`
: Set to TRUE to calculate total links per gene for hypermethylated RE DNA methylation sites with G+ links. Defaults to TRUE.

`hypomethGplusAnalysis`
: Set to TRUE to calculate total links per gene for hypomethylated RE DNA methylation sites with G+ links. Defaults to TRUE.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named "step6DNAMethylationSitesPerGeneTabulation" in its metadata containing the output of this function. This list contains `hypermethGplus` and/or `hypomethGplus` data

frames, as selected by the user, containing significant hyper- or hypomethylated G+ link counts per gene.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to tabulate both hyper- and hypomethylated G+
## RE DNA methylation site-gene links, using gene names from the input
## MultiAssayExperiment object.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Calculate linked RE DNA methylation sites per gene
returnValue <- step6DNAMethylationSitesPerGeneTabulation(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment
)

## This example is similar, but only analyzes hypomethylated RE DNA
## methylation sites.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Calculate linked RE DNA methylation sites per gene
returnValue <- step6DNAMethylationSitesPerGeneTabulation(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    hypermethGplusAnalysis = FALSE
)
```

---

step7ExpressionVsDNAMethylationScatterplots

Create scatterplots displaying the expression of the top genes and the methylation levels of each of their linked RE DNA methylation sites, optionally incorporating copy number variation, somatic mutation, and purity data

---

Description

This function takes the top genes and transcription factors by number of linked RE DNA methylation sites identified by the `step6DNAMethylationSitesPerGeneTabulation` function up to a number specified by the user, or all genes linked to selected RE DNA methylation sites specified by the user, and generates scatterplots displaying the expression level of each of these genes in the X-axis and the methylation level of each RE DNA methylation site linked to them in the Y-axis for the

hyper- and/or hypomethylated G+ analysis quadrants. The scatterplots may optionally incorporate provided copy number variation (CNV), somatic mutation (SM), and purity information for each sample.

Usage

```
step7ExpressionVsDNAMethylationScatterplots(
  TENETMultiAssayExperiment,
  geneAnnotationDataset = NA,
  hypermethGplusAnalysis = TRUE,
  hypomethGplusAnalysis = TRUE,
  topGeneNumber = 10,
  DNAMethylationSites = NA,
  simpleOrComplex = "simple",
  CNVData = NA,
  SMData = NA,
  purityData = NA,
  coreCount = 1
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step5OptimizeLinks` and `step6DNAMethylationSitesPerGeneTabulation` functions.

`geneAnnotationDataset`
: Specify a gene annotation dataset which is used to identify names for genes by their Ensembl IDs. The argument must be either a GRanges object (such as one imported via `rtracklayer::import`) or a path to a GFF3 or GTF file. Both GENCODE and Ensembl annotations are supported. Other annotation datasets may work, but have not been tested. See the "Input data" section of the vignette for information on the required dataset format. Specify NA to use the gene names listed in the "geneName" column of the elementMetadata of the rowRanges of the "expression" SummarizedExperiment object within the TENETMultiAssayExperiment object. Defaults to NA.

`hypermethGplusAnalysis`
: Set to TRUE to create scatterplots for genes with hypermethylated RE DNA methylation sites with G+ links and each of their linked RE DNA methylation sites. Defaults to TRUE.

`hypomethGplusAnalysis`
: Set to TRUE to create scatterplots for genes with hypomethylated RE DNA methylation sites with G+ links and each of their linked RE DNA methylation sites. Defaults to TRUE.

`topGeneNumber`
: Specify the number of top genes and TFs, based on the most linked RE DNA methylation sites of a given analysis type, for which to create scatterplots. Defaults to 10.

`DNAMethylationSites`
: Supply a vector of RE DNA methylation site IDs for which scatterplots will be generated, if these sites have any linked genes/TFs with expression in each specified analysis type.

`simpleOrComplex`
: Set to 'complex' to incorporate copy number variation, somatic mutation, and purity data into the scatterplots. Otherwise, set to 'simple'. If set to 'complex', copy number variation, somatic mutation, and purity data must be provided via the `CNVData`, `SMData`, and `purityData` arguments respectively. **Note:** At this time, either all or none of these optional data types must be provided. Defaults to 'simple'.

`CNVData`
: Specify a dataset containing CNV status for each of the top genes, as selected by the analysis type and 'topGeneNumber' arguments, in each sample in the TENETMultiAssayExperiment. CNV status must be an integer representing the change in copy number for each gene, with negative numbers representing a loss and positive numbers representing a gain. **Note:** Copy number changes of 2 or more will be grouped together. The dataset may be given as a data frame, matrix, or TSV file path. If it is a data frame or matrix, its rownames must contain sample names. If a TSV file is provided, the first column must contain sample names, and the first row must contain column headers. Sample names must match those in the colData of the TENETMultiAssayExperiment object. Column names must contain gene IDs followed by "_CNV". If set to NA, the data will be loaded from the colData of the TENETMultiAssayExperiment object. **Note:** If data are missing for a given gene, the plot will be generated without considering its CNV status. Defaults to NA, and is only considered if `simpleOrComplex` is set to "complex".

`SMData`
: Specify a dataset containing the somatic mutation status for each of the top genes in each sample in the TENETMultiAssayExperiment. This argument behaves the same way as the `CNVData` argument, except that the names of the columns containing SM status must end with "_SM", and the status must be an integer 0 or 1 or a string "no mutation" or "mutation". Defaults to NA.

`purityData`
: Specify the cellularity/purity data for each sample in the TENETMultiAssayExperiment. Purity values must range from 0 to 1. The dataset may be given as a vector, data frame, matrix, or TSV file path. If a vector is given, the names of the vector elements must correspond to the names of the samples in the rownames of the colData of the TENETMultiAssayExperiment object. If no names are provided for the vector, then the number of elements in the vector must equal the number of samples in the colData, and it is assumed to align with the samples as they are ordered in the colData. If a data frame, matrix, or TSV file is given, it must be in the same format as for the `CNVData` argument, except that the first column of data (excluding the rownames) must contain the purity data. If this argument is set to NA, purity data will be loaded from the "purity" column of the colData of the TENETMultiAssayExperiment object. Defaults to NA, and is only considered if 'simpleOrComplex' is set to "complex".

`coreCount`
: Argument passed as the mc.cores argument to mcmapply. See `?parallel::mcmapply` for more details. Defaults to 1.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named 'step7ExpressionVsDNAMethylationScatterplots' in its metadata with the output of this function. This list is subdivided into hypermethGplus or hypomethGplus results as selected by the user, which are further subdivided into lists with data for the top overall genes, and for top TF genes only. Each of these lists contains a final list for each of the top genes/TFs containing scatterplots for each RE DNA methylation site linked to the gene. If the user has specified RE DNA methylation sites of interest, an additional list named 'selectedDNAMethylationSites' is generated for each quadrant containing scatterplots for each gene linked to each specified RE DNA methylation site. In each scatterplot, the expression of the gene is plotted on the X-axis, and the methylation of the linked RE DNA methylation site is plotted on the Y-axis. If complex plots are being created, the CNV and SM status of each sample, if present, will be represented by each point's shape (with SM status taking precedence over CNV), and the purity of each sample will be reflected in each point's size.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to create scatterplots for the top 10
## genes and TFs by number of linked hyper- and hypomethylated RE DNA
## methylation sites, showing expression of these genes and the DNA
## methylation level of their linked RE DNA methylation sites. Gene names
## will be retrieved from the rowRanges of the 'expression'
## SummarizedExperiment object in the example MultiAssayExperiment. No CNV,
## SM, or purity data will be incorporated, and the analysis will be
## performed using one CPU core.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to create the scatterplots
returnValue <- step7ExpressionVsDNAMethylationScatterplots(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment
)

## This example demonstrates many of the analysis options, creating
## scatterplots for the top 5 genes and TFs as well as some example RE DNA
## methylation sites of interest. As before, gene names will be retrieved
## from the rowRanges of the 'expression' SummarizedExperiment object.
## Complex scatterplots are created which display each sample's CNV and SM
## status for each gene, as well as purity data, where available. The CNV,
## SM, and purity data will be taken from specific columns of the
## exampleTENETClinicalDataFrame object. The analysis will be performed using
## 8 CPU cores.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()
```

```
## Load the data frame with example clinical data for patients in the TENET
## MultiAssayExperiment object from the TENET.ExperimentHub package
exampleTENETClinicalDataFrame <-
    TENET.ExperimentHub::exampleTENETClinicalDataFrame()

## Use the example datasets to create the scatterplots
returnValue <- step7ExpressionVsDNAMethylationScatterplots(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    hypermethGplusAnalysis = FALSE,
    topGeneNumber = 5,
    DNAMethylationSites = c("cg03095778", "cg24011501", "cg12989041"),
    simpleOrComplex = "complex",
    CNVData = exampleTENETClinicalDataFrame[seq(4, 42, by = 2)],
    SMData = exampleTENETClinicalDataFrame[seq(5, 43, by = 2)],
    purityData = exampleTENETClinicalDataFrame[3],
    coreCount = 8
)
```

`step7LinkedDNAMethylationSiteCountHistograms` *Create histograms displaying the number of total genes and transcription factor genes linked to a given number of RE DNA methylation sites*

Description

This function generates histograms displaying the number of total genes and transcription factor genes linked to a given number of RE DNA methylation sites. These are designed to highlight the top overall genes and TF genes, which likely have a disproportionately large number of linked RE DNA methylation sites compared to most genes.

Usage

```
step7LinkedDNAMethylationSiteCountHistograms(
  TENETMultiAssayExperiment,
  hypermethGplusAnalysis = TRUE,
  hypomethGplusAnalysis = TRUE
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step6DNAMethylationSitesPerGeneT` function.

hypermethGplusAnalysis
: Set to TRUE to create histograms of genes linked to hypermethylated RE DNA methylation sites with G+ links. Defaults to TRUE.

hypomethGplusAnalysis
: Set to TRUE to create histograms of genes linked to hypomethylated RE DNA methylation sites with G+ links. Defaults to TRUE.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named 'step7LinkedDNAMethylationSiteCountHistograms' in its metadata, which is subdivided into `hypermethGplus` and/or `hypomethGplus` lists as selected by the user. Each of these contains histograms displaying the number of total genes and TF genes linked to a given number of RE DNA methylation sites in each of the selected analysis quadrants.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to create histograms displaying the number of
## total genes and TF genes linked to a given number of hyper- and
## hypomethylated G+ RE DNA methylation sites.
## Since we performed analyses using only TFs in the step 3 function, the
## top genes are all TFs, so a message that separate output for
## TFs will be skipped is displayed.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to create the RE DNA methylation site count
## histograms
returnValue <- step7LinkedDNAMethylationSiteCountHistograms(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment
)

## This example does the same, but only analyzes hypomethylated G+ RE DNA
## methylation sites.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to create the RE DNA methylation site count
## histogram
returnValue <- step7LinkedDNAMethylationSiteCountHistograms(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    hypermethGplusAnalysis = FALSE
)
```

step7LinkedDNAMethylationSitesMotifSearching
Search for transcription factor motifs in the vicinity of DNA methylation sites and/or within custom regions defined by the user

Description

This function takes a user-specified named list of transcription factors (TFs) and their binding motifs in the form of position weight matrices (PWMs), and/or search terms to identify additional TF binding motifs. The function identifies if each motif is found within a user-specified distance from RE DNA methylation sites in the hyper- and/or hypomethylated G+ analysis quadrants and/or sites specified by the user, and/or within specified genomic regions.

Usage

```
step7LinkedDNAMethylationSitesMotifSearching(
  TENETMultiAssayExperiment,
  hypermethGplusAnalysis = TRUE,
  hypomethGplusAnalysis = TRUE,
  DNAMethylationSites = NA,
  distanceFromREDNAMethylationSites = 100,
  GRangesToSearch = NA,
  andStrings = NULL,
  orStrings = NULL,
  notStrings = NULL,
  TFMotifList,
  useOnlyDNAMethylationSitesLinkedToTFs = TRUE,
  geneAnnotationDataset = NA,
  DNAMethylationArray = NA,
  matchPWMMinScore = "75%",
  coreCount = 1
)
```

Arguments

TENETMultiAssayExperiment
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step5OptimizeLinks` function if `hypermethGplusAnalysis` or `hypomethGplusAnalysis` are TRUE.

hypermethGplusAnalysis
: Set to TRUE to search for motifs in the vicinity of hypermethylated RE DNA methylation sites with at least one linked TF. **Note**: If `useOnlyDNAMethylationSitesLinkedToTFs` is also TRUE, only RE DNA methylation sites linked to TFs specified via the `TFMotifList` argument will be used. Defaults to TRUE.

hypomethGplusAnalysis
: Set to TRUE to search for motifs in the vicinity of hypomethylated RE DNA methylation sites with at least one linked TF. **Note**: If useOnlyDNAMethylationSitesLinkedToTFs is also TRUE, only RE DNA methylation sites linked to TFs specified via the TFMotifList argument will be used. Defaults to TRUE.

DNAMethylationSites
: Supply a vector of IDs of DNA methylation sites to search for motifs in the vicinity of these sites, in addition to any RE DNA methylation sites selected by the hypermethGplusAnalysis and hypomethGplusAnalysis arguments. If set to NA, no additional DNA methylation sites will be included in the search. Defaults to NA.

distanceFromREDNAMethylationSites
: Specify the positive integer distance from the DNA methylation sites selected by the hypermethGplusAnalysis, hypomethGplusAnalysis, and DNAMethylationSites arguments within which motif searching will be performed. Defaults to 100.

GRangesToSearch
: Specify a GRanges object which contains genomic coordinates of regions within which to search for motifs. The coordinates should correspond to the human hg38 genome. Any regions included in this GRanges object will be combined with regions defined by the hypermethGplusAnalysis, hypomethGplusAnalysis, DNAMethylationSites, and distanceFromREDNAMethylationSites arguments. If set to NA, no additional regions will be included in the motif search. Defaults to NA.

andStrings
: Specify a vector of values which will be provided to the andStrings argument of the query() function in the MotifDb package, used to search for motif PWMs. Potential values include species and transcription factor database names to refine the search. Set to NULL to include no terms in this search. Defaults to NULL. **Note:** If both andStrings and orStrings are set to NULL, only the PWMs specified by the TFMotifList argument will be used.

orStrings
: Specify a vector of values which will be provided to the orStrings argument of the query() function in the MotifDb package, used to search for motif PWMs. Potential values include names of specific TFs to limit the search to. The value "humanTranscriptionFactors" may be specified to use all TFs identified in 'The Human Transcription Factors' by Lambert et al. 2018. Set to NULL to include no terms in this search. Defaults to NULL. **Note:** If both andStrings and orStrings are set to NULL, only the PWMs specified by the TFMotifList argument will be used.

notStrings
: Specify a vector of values which will be provided to the notStrings argument of the query() function in the MotifDb package, used to exclude results from the motif PWM search. The value "humanTranscriptionFactors" may be specified to use all TFs identified in 'The Human Transcription Factors' by Lambert et al. 2018. Set to NULL to exclude no terms from this search. Defaults to NULL.

TFMotifList
: Specify a named list mapping transcription factor gene names and/or IDs to their respective motif position weight matrix (PWM). The PWMs should be in the form of a 4xN matrix. PWMs specified in this list are combined with any TF PWMs retrieved via the MotifDb package using the andStrings, orStrings,

and notStrings arguments. Set to NA to only include PWMs retrieved by the MotifDb package in the search.

useOnlyDNAMethylationSitesLinkedToTFs

If set to TRUE, only hypomethylated or hypermethylated RE DNA methylation sites, as selected by the hypermethGplusAnalysis and hypomethGplusAnalysis arguments, which are found to be linked to the TFs in the given TFMotifList by TENET will be analyzed. To use this functionality, at least one of hypermethGplusAnalysis or hypomethGplusAnalysis must be set to TRUE, DNAMethylationSites, andStrings, and orStrings must be NA, and the name of each PWM in the list given to TFMotifList must match the gene name or Ensembl ID of a gene in the TENETMultiAssayExperiment with RE DNA methylation sites linked to it for the specified analysis types. Defaults to TRUE.

geneAnnotationDataset

Specify a gene annotation dataset which is used to identify names for genes by their Ensembl IDs. The argument must be either a GRanges object (such as one imported via rtracklayer::import) or a path to a GFF3 or GTF file. Both GENCODE and Ensembl annotations are supported. Other annotation datasets may work, but have not been tested. See the "Input data" section of the vignette for information on the required dataset format. Specify NA to use the gene names listed in the "geneName" column of the elementMetadata of the rowRanges of the "expression" SummarizedExperiment object within the TENETMultiAssayExperiment object. Defaults to NA.

DNAMethylationArray

Specify the name of a DNA methylation probe array supported by the sesameData package (see ?sesameData::sesameData_getManifestGRanges). If an array is specified, RE DNA methylation sites and their locations in that array's manifest are cross-referenced with RE DNA methylation site IDs included in the rownames of the methylation dataset provided in the "methylation" SummarizedExperiment object within the TENETMultiAssayExperiment object, and only those overlapping will be considered for analysis. If set to NA, all RE DNA methylation sites with locations listed in the rowRanges of the "methylation" SummarizedExperiment object are used. Defaults to NA.

matchPWMMinScore

Specify the min.score argument passed to the matchPWM function for motif searching. See ?Biostrings::matchPWM for more details. Defaults to "75%".

coreCount Argument passed as the mc.cores argument to mclapply. See ?parallel::mclapply for more details. Defaults to 1.

Details

Note: Using many input motifs or RE DNA methylation sites may cause the search to take a significant amount of time, so in this case, using multiple CPU cores is highly recommended.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named 'step7LinkedDNAMethylationSitesMotifSearching' in its metadata containing the output of this function. This list includes the object "DNAMethylationSitesGRanges"

containing the regions in which motif searching was performed, "TFMotifPWMList" containing the TF PWMs searched for, "TFMotifSeqLogoList" which includes visual sequence logo representations of these PWMs, the "DNAMethylationSitesMotifOccurrences" data frame, which notes the location and PWM of all motifs found, the regions they were found within, as well as a "totalMotifOccurrencesPerDNAMethylationSite" data frame noting how many times each PWM listed in the "TFMotifPWMList" was found in each region in the "DNAMethylationSitesGRanges" object. If `useOnlyDNAMethylationSitesLinkedToTFs` was set to TRUE, an additional data frame "linkedUniqueDNAMethylationSitesTFOverlap" is included, which notes which TFs in the "TFMotifPWMList" the hyper- or hypomethylated RE DNA methylation sites used in the analysis were linked to; otherwise, it will be NA.

Examples

```
## Show available motifs for example TF FOXA1
names(MotifDb::query(MotifDb::MotifDb, "FOXA1"))

## The sequence logos for all input motifs will be included in the output
## of this function. Alternatively, individual motifs can be visualized
## with the seqLogo function from the seqLogo package.
seqLogo::seqLogo(MotifDb::query(MotifDb::MotifDb, "FOXA1")[[3]])

## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to perform motif searching in the vicinity of
## all hyper- and hypomethylated RE DNA methylation sites linked to the
## FOXA1 and ESR1 TF genes. The motifs these TFs bind to will be retrieved
## via the MotifDb package. Gene names and locations, and the locations of RE
## DNA methylation sites, will be retrieved from the rowRanges of the
## 'expression' and 'methylation' SummarizedExperiment objects in the
## example MultiAssayExperiment. Regions within 100 bp of linked RE DNA
## methylation sites will be considered in the search, and a motif similarity
## threshold of 75% will be used. The analysis will be performed using one
## CPU core.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to perform the motif searching
returnValue <- step7LinkedDNAMethylationSitesMotifSearching(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    orStrings = c("FOXA1", "ESR1")
)

## This example is similar, but performs motif searching in the vicinity
## of only hypomethylated RE DNA methylation sites linked to the FOXA1 and
## ESR1 TF genes. Regions within 50 bp of linked RE DNA methylation sites
## will be considered in the search, and a motif similarity threshold of 80%
## will be used. The analysis will be performed using 8 CPU cores.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
```

```
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to perform the motif searching
returnValue <- step7LinkedDNAMethylationSitesMotifSearching(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    orStrings = c("FOXA1", "ESR1"),
    hypermethGplusAnalysis = FALSE,
    distanceFromREDNAMethylationSites = 50,
    matchPWMMinScore = "80%",
    coreCount = 8
)

## This example demonstrates how to search for motifs in the vicinity of only
## specific DNA methylation sites, regardless of whether they are linked to
## TFs, and how to specify custom motif position weight matrices (PWMs),
## while also including motifs for all human transcription factors in the
## SwissRegulon database accessed by the `MotifDb::query()` function. The
## rest of the options are set to the default values described in the first
## example above.

## Create a list of example PWMs. For the purposes of this example, they
## are retrieved using the MotifDb package, although this functionality is
## intended for user-specified motifs that do not appear in the MotifDb
## database.
exampleTFMotifList <- list(
    "FOXA1" = MotifDb::query(MotifDb::MotifDb, "FOXA1")[[3]],
    "MYBL2" = MotifDb::query(MotifDb::MotifDb, "MYBL2")[[5]]
)

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to perform the motif searching
returnValue <- step7LinkedDNAMethylationSitesMotifSearching(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    hypermethGplusAnalysis = FALSE,
    hypomethGplusAnalysis = FALSE,
    DNAMethylationSites = c("cg04134755", "cg10216151"),
    andStrings = c("Hsapiens", "SwissRegulon"),
    orStrings = "humanTranscriptionFactors",
    TFMotifList = exampleTFMotifList,
    useOnlyDNAMethylationSitesLinkedToTFs = FALSE
)
```

`step7SelectedDNAMethylationSitesCaseVsControlBoxplots`
Generate boxplots or violin plots comparing the methylation level of the specified RE DNA methylation sites in case and control samples

---

Description

This function takes a vector of RE DNA methylation sites specified by the user and generates boxplots or violin plots displaying the methylation level of each of these DNA methylation sites in the case compared to control samples, along with the results of a Student's t-test comparing the methylation level between these two groups.

Usage

```
step7SelectedDNAMethylationSitesCaseVsControlBoxplots(
  TENETMultiAssayExperiment,
  DNAMethylationSites,
  violinPlots = FALSE,
  coreCount = 1
)
```

Arguments

`TENETMultiAssayExperiment`	Specify a MultiAssayExperiment object containing a methylation SummarizedExperiment object, such as one created by the TCGADownloader function.
`DNAMethylationSites`	Supply a vector of RE DNA methylation site IDs for which to create boxplots or violin plots with the methylation of those RE DNA methylation sites.
`violinPlots`	Set to TRUE to generate violin plots instead of boxplots. Defaults to FALSE.
`coreCount`	Argument passed as the mc.cores argument to mcmapply. See `?parallel::mcmapply` for more details. Defaults to 1.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named 'step7SelectedDNAMethylationSitesCaseVsControlBoxplots' in its metadata, which contains boxplots or violin plots comparing the methylation of the RE DNA methylation sites of interest in the case and control samples. The titles of the plots contain the ID of the RE DNA methylation site and the Student's t-test p-value.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to generate boxplots for several selected
## RE DNA methylation sites, using one CPU core.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
```

```
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to create the boxplots
returnValue <- step7SelectedDNAMethylationSitesCaseVsControlBoxplots(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    DNAMethylationSites = c("cg03095778", "cg24011501", "cg12989041"),
    coreCount = 1
)
```

`step7StatesForLinks` *Identify which of the case samples harbor each of the identified regulatory element DNA methylation site-gene links*

Description

This function generates data frames indicating whether each of the case samples harbors each RE DNA methylation site-gene link from the hyper- and/or hypomethylated G+ analysis quadrants. A given sample harbors a link if the RE DNA methylation site in the link is hyper- and/or hypomethylated in that sample, and if expression of the gene in the link in that sample is significantly less than or greater than, respectively, the mean expression of the gene in the control samples, as determined by a Bonferroni-corrected 1-sided t-test with a p-value threshold of 0.05. This may aid in identifying case samples for further analyses.

Usage

```
step7StatesForLinks(
  TENETMultiAssayExperiment,
  hypermethGplusAnalysis = TRUE,
  hypomethGplusAnalysis = TRUE,
  coreCount = 1
)
```

Arguments

`TENETMultiAssayExperiment`	Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step2GetDifferentiallyMethylated` and `step5OptimizeLinks` functions.
`hypermethGplusAnalysis`	Set to TRUE to analyze which case samples harbor each hypermethylated G+ RE DNA methylation site-gene link. Defaults to TRUE.
`hypomethGplusAnalysis`	Set to TRUE to analyze which case samples harbor each hypomethylated G+ RE DNA methylation site-gene link. Defaults to TRUE.
`coreCount`	Argument passed as the mc.cores argument to mcmapply. See `?parallel::mcmapply` for more details. Defaults to 1.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named 'step7StatesForLinks' in its metadata containing the output of this function. This list contains `hypermethGplus` and/or `hypomethGplus` data frames, as selected by the user, which contain columns for each case sample and rows for each RE DNA methylation site-gene link for that analysis type. A 1 at the intersection of a row and column indicates the sample harbors that link and a 0 indicates it does not; an NA value indicates that the sample lacks methylation data for the site or expression data for the gene.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to create data frames for link states between
## genes and hyper- and hypomethylated G+ RE DNA methylation sites. The
## analysis will be performed using one CPU core.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to perform the analysis
returnValue <- step7StatesForLinks(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment
)

## This example is similar, but it analyzes only hypomethylated G+ RE DNA
## methylation sites, and the analysis is performed using 8 CPU cores.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to perform the analysis
returnValue <- step7StatesForLinks(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    hypermethGplusAnalysis = FALSE,
    coreCount = 8
)
```

---

step7TopGenesCaseVsControlExpressionBoxplots

Generate boxplots or violin plots comparing the expression level of the top genes and transcription factors in case and control samples

---

Description

This function takes the top genes and transcription factors (TFs) for each analysis type by number of linked RE DNA methylation sites identified by the `step6DNAMethylationSitesPerGeneTabulation` function, up to the number specified by the user, and generates boxplots or violin plots displaying the expression level of each of these genes in the case compared to control samples, along with the results of a Student's t-test comparing the expression level between these two groups.

Usage

```
step7TopGenesCaseVsControlExpressionBoxplots(
  TENETMultiAssayExperiment,
  geneAnnotationDataset = NA,
  hypermethGplusAnalysis = TRUE,
  hypomethGplusAnalysis = TRUE,
  topGeneNumber = 10,
  violinPlots = FALSE,
  coreCount = 1
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step5OptimizeLinks` and `step6DNAMethylationSitesPerGeneTabulation` functions.

`geneAnnotationDataset`
: Specify a gene annotation dataset which is used to identify names for genes by their Ensembl IDs. The argument must be either a GRanges object (such as one imported via `rtracklayer::import`) or a path to a GFF3 or GTF file. Both GENCODE and Ensembl annotations are supported. Other annotation datasets may work, but have not been tested. See the "Input data" section of the vignette for information on the required dataset format. Specify NA to use the gene names listed in the "geneName" column of the elementMetadata of the rowRanges of the "expression" SummarizedExperiment object within the TENETMultiAssayExperiment object. Defaults to NA.

`hypermethGplusAnalysis`
: Set to TRUE to create plots for the top genes and TFs with the most hypermethylated RE DNA methylation sites with G+ links. Defaults to TRUE.

`hypomethGplusAnalysis`
: Set to TRUE to create expression boxplots or violin plots for the top genes and TFs with the most hypomethylated RE DNA methylation sites with G+ links. Defaults to TRUE.

`topGeneNumber` Specify the number of top genes and TFs, based on the most linked RE DNA methylation sites of a given analysis type, for which to generate expression boxplots or violin plots. Defaults to 10.

`violinPlots` Set to TRUE to generate violin plots instead of boxplots. Defaults to FALSE.

coreCount Argument passed as the mc.cores argument to mcmapply. See ?parallel::mcmapply for more details. Defaults to 1.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named 'step7TopGenesCaseVsControlExpressionBoxplots' in its metadata containing the output of this function. This list contains hypermethGplus and/or hypomethGplus lists, as selected by the user, which contain lists for the top overall genes and the top TF genes. These lists contain boxplots or violin plots showing the expression of the gene of interest in the case and control samples, with Student's t-test p-values and the name and ID of the gene in the title.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to create boxplots comparing expression in
## case and control samples for the top 10 genes and TFs by number of linked
## hyper- and hypomethylated RE DNA methylation sites. Gene names will be
## retrieved from the rowRanges of the 'expression' SummarizedExperiment
## object in the example MultiAssayExperiment. The analysis will be performed
## using one CPU core.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to create the boxplots
returnValue <- step7TopGenesCaseVsControlExpressionBoxplots(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment
)

## This example is similar, but it only creates boxplots for the top 5 genes
## and TFs by number of linked hypomethylated RE DNA methylation sites, and
## the analysis is performed using 8 CPU cores.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to create the boxplots
returnValue <- step7TopGenesCaseVsControlExpressionBoxplots(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    hypermethGplusAnalysis = FALSE,
    topGeneNumber = 5,
    coreCount = 8
)
```

`step7TopGenesCircosPlots` *Generate Circos plots displaying the links between the top identified genes and each of the RE DNA methylation sites linked to them*

Description

This function takes the top genes and TFs by number of linked regulatory element DNA methylation sites identified by the `step6DNAMethylationSitesPerGeneTabulation` function, up to the number specified by the user, and generates Circos plots for each gene showing the genomic links between each gene and each RE DNA methylation site linked to the gene for the analysis types specified.

Usage

```
step7TopGenesCircosPlots(
  TENETMultiAssayExperiment,
  DNAMethylationArray = NA,
  geneAnnotationDataset = NA,
  hypermethGplusAnalysis = TRUE,
  hypomethGplusAnalysis = TRUE,
  topGeneNumber = 10,
  coreCount = 1
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step5OptimizeLinks` and `step6DNAMethylationSitesPerGeneTabulation` functions.

`DNAMethylationArray`
: Specify the name of a DNA methylation probe array supported by the sesameData package (see `?sesameData::sesameData_getManifestGRanges`). If an array is specified, RE DNA methylation sites and their locations in that array's manifest are cross-referenced with RE DNA methylation site IDs included in the rownames of the methylation dataset provided in the "methylation" SummarizedExperiment object within the TENETMultiAssayExperiment object, and only those overlapping will be considered for analysis. If set to NA, all RE DNA methylation sites with locations listed in the rowRanges of the "methylation" SummarizedExperiment object are used. Defaults to NA.

`geneAnnotationDataset`
: Specify a gene annotation dataset which is used to identify names for genes by their Ensembl IDs. The argument must be either a GRanges object (such as one imported via `rtracklayer::import`) or a path to a GFF3 or GTF file. Both GENCODE and Ensembl annotations are supported. Other annotation

datasets may work, but have not been tested. See the "Input data" section of the vignette for information on the required dataset format. Specify NA to use the gene names listed in the "geneName" column of the elementMetadata of the rowRanges of the "expression" SummarizedExperiment object within the TENETMultiAssayExperiment object. Defaults to NA.

`hypermethGplusAnalysis`
: Set to TRUE to create Circos plots displaying genomic links between the top genes and TFs by most hypermethylated RE DNA methylation sites with G+ links and their linked RE DNA methylation sites of that type. Defaults to TRUE.

`hypomethGplusAnalysis`
: Set to TRUE to create Circos plots displaying genomic links between the top genes and TFs by most hypomethylated RE DNA methylation sites with G+ links and their linked RE DNA methylation sites of that type. Defaults to TRUE.

`topGeneNumber`
: Specify the number of top genes and TFs, by number of linked RE DNA methylation sites of a given analysis type, for which to generate Circos plots showing genomic links between the genes and each of their linked RE DNA methylation sites. Defaults to 10.

`coreCount`
: Argument passed as the mc.cores argument to mclapply. See `?parallel::mclapply` for more details. Defaults to 1.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named 'step7TopGenesCircosPlots' in its metadata containing the output of this function. This list contains `hypermethGplus` and/or `hypomethGplus` lists, as selected by the user, which contain lists for the top overall genes and the top TF genes. These lists contain Circos plots visualizing the genomic links between each gene and its linked RE DNA methylation sites for the selected analysis type.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to create Circos plots for the top 10
## genes and TFs by number of linked hyper- and hypomethylated RE DNA
## methylation sites. Gene names and locations and RE DNA methylation site
## locations will be retrieved from the rowRanges of the 'expression' and
## 'methylation' SummarizedExperiment objects in the example
## MultiAssayExperiment. The analysis will be performed using one CPU core.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to create Circos plots
returnValue <- step7TopGenesCircosPlots(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment
)

## This example is similar, but creates Circos plots for only the top 5 genes
```

```
## and TFs by number of linked hypomethylated RE DNA methylation sites.
## RE DNA methylation site IDs and locations are retrieved from the
## HM450 array via the sesameData package, and eight CPU cores are used.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to create Circos plots
returnValue <- step7TopGenesCircosPlots(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    DNAMethylationArray = "HM450",
    hypermethGplusAnalysis = FALSE,
    topGeneNumber = 5,
    coreCount = 8
)
```

`step7TopGenesDNAMethylationHeatmaps`

Generate heatmaps displaying the methylation level of all RE DNA methylation sites linked to the top genes and transcription factors, along with the expression of those genes in the column headers, in the case samples within the supplied MultiAssayExperiment object

---

Description

This function takes the top genes and transcription factors (TFs) for each analysis type by number of linked RE DNA methylation sites identified by the `step6DNAMethylationSitesPerGeneTabulation` function, up to the number specified by the user, and generates heatmaps displaying the methylation level of the unique RE DNA methylation sites linked to any of those genes, along with the expression of those genes in the case samples only.

Usage

```
step7TopGenesDNAMethylationHeatmaps(
  TENETMultiAssayExperiment,
  geneAnnotationDataset = NA,
  hypermethGplusAnalysis = TRUE,
  hypomethGplusAnalysis = TRUE,
  topGeneNumber = 10
)
```

Arguments

`TENETMultiAssayExperiment`

Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must also contain the results from the

step2GetDifferentiallyMethylatedSites, step5OptimizeLinks, and step6DNAMethylationSitesP functions.

geneAnnotationDataset
: Specify a gene annotation dataset which is used to identify names for genes by their Ensembl IDs. The argument must be either a GRanges object (such as one imported via rtracklayer::import) or a path to a GFF3 or GTF file. Both GENCODE and Ensembl annotations are supported. Other annotation datasets may work, but have not been tested. See the "Input data" section of the vignette for information on the required dataset format. Specify NA to use the gene names listed in the "geneName" column of the elementMetadata of the rowRanges of the "expression" SummarizedExperiment object within the TENETMultiAssayExperiment object. Defaults to NA.

hypermethGplusAnalysis
: Set to TRUE to create heatmaps showing the methylation levels of RE DNA methylation sites linked to the top genes and TFs with the most hypermethylated RE DNA methylation sites with G+ links. Defaults to TRUE.

hypomethGplusAnalysis
: Set to TRUE to create heatmaps showing the methylation levels of RE DNA methylation sites linked to the top genes and TFs with the most hypomethylated RE DNA methylation sites with G+ links. Defaults to TRUE.

topGeneNumber
: Specify the number of top genes and TFs, based on the most linked RE DNA methylation sites of a given analysis type, for which to generate heatmaps with their linked RE DNA methylation sites' methylation levels. Defaults to 10.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named 'step7TopGenesDNAMethylationHeatmaps' in its metadata containing the output of this function. This list contains hypermethGplus and/or hypomethGplus lists, as selected by the user, which contain heatmaps for the top overall genes and the top TF genes. These heatmaps show the expression of the top genes/TFs in the column headers and the methylation of their unique linked RE DNA methylation sites in the body. Column dendrograms are included to identify subsets of the case samples which display particular expression or methylation patterns in the top genes and their linked RE DNA methylation sites.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to create heatmaps for the top 10 genes and
## TFs by number of linked hyper- and hypomethylated RE DNA methylation
## sites and the unique RE DNA methylation sites linked to those genes.
## Gene names will be retrieved from the rowRanges of the 'expression'
## SummarizedExperiment object in the example MultiAssayExperiment.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()
```

```
## Use the example dataset to create methylation heatmaps
returnValue <- step7TopGenesDNAMethylationHeatmaps(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment
)

## This example is similar, but creates heatmaps for only the top 5 genes
## and TFs by number of linked hypomethylated RE DNA methylation sites.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to create methylation heatmaps
returnValue <- step7TopGenesDNAMethylationHeatmaps(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    hypermethGplusAnalysis = FALSE,
    topGeneNumber = 5
)
```

`step7TopGenesExpressionCorrelationHeatmaps`
Generate mirrored heatmaps displaying the correlation of the expression values of the top genes and TFs

---

Description

This function takes the top genes and TFs for each analysis type by number of linked RE DNA methylation sites identified by the `step6DNAMethylationSitesPerGeneTabulation` function, up to the number specified by the user, and generates heatmaps displaying the correlation of the expression of each of the top genes and TFs in the case samples. Each of the genes is displayed in both the rows and columns, so the heatmaps are mirrored, with correlation values of each gene to itself displayed in a diagonal line in the center of the heatmaps. Red values represent positive correlation and blue values represent negative correlation, with darker colors representing a stronger correlation. Dendrograms are included to identify genes which are closely related in expression correlation.

Usage

```
step7TopGenesExpressionCorrelationHeatmaps(
  TENETMultiAssayExperiment,
  geneAnnotationDataset = NA,
  hypermethGplusAnalysis = TRUE,
  hypomethGplusAnalysis = TRUE,
  topGeneNumber = 10
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step6DNAMethylationSitesPerGeneT` function.

`geneAnnotationDataset`
: Specify a gene annotation dataset which is used to identify names for genes by their Ensembl IDs. The argument must be either a GRanges object (such as one imported via `rtracklayer::import`) or a path to a GFF3 or GTF file. Both GENCODE and Ensembl annotations are supported. Other annotation datasets may work, but have not been tested. See the "Input data" section of the vignette for information on the required dataset format. Specify NA to use the gene names listed in the "geneName" column of the elementMetadata of the rowRanges of the "expression" SummarizedExperiment object within the TENETMultiAssayExperiment object. Defaults to NA.

`hypermethGplusAnalysis`
: Set to TRUE to create heatmaps and tables showing expression correlation values for the top genes and TFs with the most hypermethylated RE DNA methylation sites with G+ links. Defaults to TRUE.

`hypomethGplusAnalysis`
: Set to TRUE to create heatmaps and tables showing expression correlation values for the top genes and TFs with the most

`topGeneNumber`
: Specify the number of top genes and TFs, based on the most linked RE DNA methylation sites of a given analysis type, for which to generate expression correlation heatmaps and tables. Defaults to 10.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named 'step7TopGenesExpressionCorrelationHeatmaps' in its metadata containing the output of this function. This list contains `hypermethGplus` and/or `hypomethGplus` lists, as selected by the user, which contain lists for the top overall genes and top TF genes. These lists contain a mirrored heatmap displaying the expression correlation values for these genes and a data frame containing the names and correlation values for each gene.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to create correlation heatmaps, and
## corresponding tables, for the top 10 genes and TFs by number of
## linked hyper- and hypomethylated RE DNA methylation sites. Gene names will
## be retrieved from the rowRanges of the 'expression' SummarizedExperiment
## object in the example MultiAssayExperiment.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()
```

```
## Use the example dataset to create expression correlation heatmaps
returnValue <- step7TopGenesExpressionCorrelationHeatmaps(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment
)

## This example is similar, but creates heatmaps and tables for only the
## top 5 genes and TFs by number of linked hypomethylated RE DNA methylation
## sites.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to create expression correlation heatmaps
returnValue <- step7TopGenesExpressionCorrelationHeatmaps(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    hypermethGplusAnalysis = FALSE,
    topGeneNumber = 5
)
```

`step7TopGenesOverlappingLinkedDNAMethylationSitesHeatmaps`
Generate binary heatmaps displaying which of the top genes and transcription factors share links with each of the unique regulatory element DNA methylation sites linked to at least one top gene/TF

Description

This function takes the top genes and TFs for each analysis type by number of linked RE DNA methylation sites identified by the `step6DNAMethylationSitesPerGeneTabulation` function, up to the number specified by the user, and identifies the unique RE DNA methylation sites linked to them, then generates two-color binary heatmaps displaying which of the top genes and TFs the RE DNA methylation sites are linked to, as well as data frames with that information.

Usage

```
step7TopGenesOverlappingLinkedDNAMethylationSitesHeatmaps(
  TENETMultiAssayExperiment,
  geneAnnotationDataset = NA,
  hypermethGplusAnalysis = TRUE,
  hypomethGplusAnalysis = TRUE,
  topGeneNumber = 10
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step5OptimizeLinks` and `step6DNAMethylationSitesPerGeneTabulation` functions.

`geneAnnotationDataset`
: Specify a gene annotation dataset which is used to identify names for genes by their Ensembl IDs. The argument must be either a GRanges object (such as one imported via `rtracklayer::import`) or a path to a GFF3 or GTF file. Both GENCODE and Ensembl annotations are supported. Other annotation datasets may work, but have not been tested. See the "Input data" section of the vignette for information on the required dataset format. Specify NA to use the gene names listed in the "geneName" column of the elementMetadata of the rowRanges of the "expression" SummarizedExperiment object within the TENETMultiAssayExperiment object. Defaults to NA.

`hypermethGplusAnalysis`
: Set to TRUE to create heatmaps and tables showing the linked RE DNA methylation sites for the top genes and TFs with the most hypermethylated RE DNA methylation sites with G+ links. Defaults to TRUE.

`hypomethGplusAnalysis`
: Set to TRUE to create heatmaps and tables showing the linked RE DNA methylation sites for the top genes and TFs with the most hypomethylated RE DNA methylation sites with G+ links. Defaults to TRUE.

`topGeneNumber`
: Specify the number of top genes and TFs, based on the most linked RE DNA methylation sites of a given analysis type, for which to generate linked RE DNA methylation site heatmaps and tables. Defaults to 10.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named 'step7TopGenesOverlappingLinkedDNAMethylationSitesHeatmaps' in its metadata containing the output of this function. This list contains `hypermethGplus` and/or `hypomethGplus` lists, as selected by the user, which contain lists for the top overall genes and top TF genes. These lists contain binary heatmaps displaying the top genes/TFs in the columns and the unique RE DNA methylation sites linked to these genes in the rows, with black indicating that the given RE DNA methylation site is linked to the given gene. Dendrograms are included to identify blocks of RE DNA methylation sites that are linked to similar genes. Data frames are also included which represent the links numerically, with 1s indicating a link is present.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to create overlap heatmaps, and corresponding
## data frames, for the top 10 genes and TFs by number of linked hyper- and
## hypomethylated RE DNA methylation sites. Gene names will be retrieved
## from the rowRanges of the 'expression' SummarizedExperiment object in the
## example MultiAssayExperiment.
```

```
## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to create overlap heatmaps
returnValue <- step7TopGenesOverlappingLinkedDNAMethylationSitesHeatmaps(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment
)

## This example is similar, but creates overlap heatmaps and corresponding
## data frames for only the top 5 genes and TFs by number of linked
## hypomethylated RE DNA methylation sites.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to create overlap heatmaps
returnValue <- step7TopGenesOverlappingLinkedDNAMethylationSitesHeatmaps(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    hypermethGplusAnalysis = FALSE,
    topGeneNumber = 5
)
```

`step7TopGenesSurvival` *Perform Kaplan-Meier and Cox regression analyses to assess the association of patient survival with the expression of top genes and transcription factors and methylation of their linked RE DNA methylation sites*

Description

This function takes the top genes and transcription factors (TFs) by number of linked RE DNA methylation sites identified by the `step6DNAMethylationSitesPerGeneTabulation` function, up to the number specified by the user, along with patient survival data, and generates plots and tables with statistics assessing the association of patient survival with the expression of top genes and transcription factors and methylation of their linked RE DNA methylation sites, using groupings based on percentile cutoffs or Jenks natural breaks for Kaplan-Meier analyses.

Usage

```
step7TopGenesSurvival(
  TENETMultiAssayExperiment,
  geneAnnotationDataset = NA,
  hypermethGplusAnalysis = TRUE,
```

```
  hypomethGplusAnalysis = TRUE,
  topGeneNumber = 10,
  vitalStatusData = NA,
  survivalTimeData = NA,
  highProportion = 0.5,
  lowProportion = 0.5,
  survivalGroupingCutoffs = NA,
  jenksBreaksGroupCount = NA,
  generatePlots = TRUE,
  coreCount = 1
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step2GetDifferentiallyMethylated`, `step5OptimizeLinks`, and `step6DNAMethylationSitesPerGeneTabulation` functions. The object's colData must contain 'vital_status' and 'time' columns containing data on the patients' survival status and time to event/censorship, respectively.

`geneAnnotationDataset`
: Specify a gene annotation dataset which is used to identify names for genes by their Ensembl IDs. The argument must be either a GRanges object (such as one imported via `rtracklayer::import`) or a path to a GFF3 or GTF file. Both GENCODE and Ensembl annotations are supported. Other annotation datasets may work, but have not been tested. See the "Input data" section of the vignette for information on the required dataset format. Specify NA to use the gene names listed in the "geneName" column of the elementMetadata of the rowRanges of the "expression" SummarizedExperiment object within the TENETMultiAssayExperiment object. Defaults to NA.

`hypermethGplusAnalysis`
: Set to TRUE to perform survival analyses on the top genes and TFs by most hypermethylated RE DNA methylation sites with G+ links, as well as their linked RE DNA methylation sites.

`hypomethGplusAnalysis`
: Set to TRUE to perform survival analyses on the top genes and TFs by most hypomethylated RE DNA methylation sites with G+ links, as well as their linked RE DNA methylation sites.

`topGeneNumber` Specify the number of top genes and TFs, based on the most linked RE DNA methylation sites of a given analysis type, for which to perform survival analyses. Defaults to 10.

`vitalStatusData`
: Specify the patient vital status data for samples in the TENETMultiAssayExperiment. Vital status should be given in the form of either "alive" or "dead" (case-insensitive), or 1 or 2, indicating that the sample was collected from a

patient who was alive/censored or dead/reached the outcome of interest, respectively. These data can be given as a vector, data frame, matrix, or path to a TSV file. Given sample names must match the names of the samples in the colData of the TENETMultiAssayExperiment. If a vector is given, the names of its elements must be the sample names; if it has no names, its length must equal the number of samples in the colData, and its values must be in the same order as the samples in the colData. If a data frame or matrix is given, its rownames must contain the sample names, and its first column must contain the vital status. If a TSV file is given, its first column must contain the sample names, its second column must contain the vital status, and its first row must contain column names. If set to NA, vital status data will be retrieved from the "vital_status" column of the colData of the TENETMultiAssayExperiment. Defaults to NA.

`survivalTimeData`
: Specify the numeric survival time data for samples in the TENETMultiAssayExperiment. These data can be given as a vector, data frame, matrix, or path to a TSV file; see the documentation for `vitalStatusData` for more information. If set to NA, survival time data will be retrieved from the "time" column of the colData of the TENETMultiAssayExperiment. Defaults to NA.

`highProportion`
: Specify the proportion of all samples to include in the high expression/methylation group for Kaplan-Meier survival analyses as a number ranging from 0 to 1. **Note:** If the `survivalGroupingCutoffs` or `jenksBreaksGroupCount` argument is specified, this argument will be ignored. Defaults to 0.5.

`lowProportion`
: Specify the proportion of all samples to include in the low expression/methylation group for Kaplan-Meier survival analyses as a number ranging from 0 to 1. **Note:** If the `survivalGroupingCutoffs` or `jenksBreaksGroupCount` argument is specified, this argument will be ignored. If both `lowProportion` and `highProportion` are set to 0.5, samples at exactly the 50th percentile will be assigned to the "Low" group. Defaults to 0.5.

`survivalGroupingCutoffs`
: To use custom sample grouping, specify a data frame or matrix with two columns and *n* rows, where *n* is the number of groups the samples should be broken into, and values ranging from 0 to 1 reflecting the proportion of samples to include in each group. Values in the first column should reflect the minimum proportion, and values in the second column should reflect the maximum proportion (non-inclusive if not 1). If the object has row names, they will be used to name the groups. If specified, the `highProportion` and `lowProportion` arguments will be ignored. Defaults to NA.

`jenksBreaksGroupCount`
: Specify the number of groups into which to break the survival data as a positive integer. Cutoffs for each group will be generated using Jenks natural breaks optimization. If specified, the `highProportion` and `lowProportion` arguments will be ignored. Defaults to NA.

`generatePlots`
: Set to TRUE to generate plots displaying the Kaplan-Meier survival results for the top genes and TFs of interest and their linked RE DNA methylation sites. Defaults to TRUE.

`coreCount`
: Argument passed as the mc.cores argument to mclapply. See `?parallel::mclapply` for more details. Defaults to 1.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named 'step7TopGenesSurvival' in its metadata containing the output of this function. This list contains `hypermethGplus` and/or `hypomethGplus` lists, as selected by the user, which contain lists for the top overall genes and top TF genes. Each contains a list of data frames containing survival statistics for the top genes/TFs and their linked RE DNA methylation sites from both Kaplan-Meier and Cox regression analyses, and a list of Kaplan-Meier plots if `generatePlots` is TRUE.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to perform Kaplan-Meier and Cox regression
## survival analyses on the top 10 genes and TFs by number of linked hyper-
## and hypomethylated RE DNA methylation sites, and on all unique RE DNA
## methylation sites linked to those genes. The vital status and
## survival time of patients will be taken from the "vital_status" and "time"
## columns of the colData of the example MultiAssayExperiment. Gene names
## will be retrieved from the rowRanges of the 'expression'
## SummarizedExperiment object in the example MultiAssayExperiment. In the
## Kaplan-Meier analyses, the patient samples with complete clinical
## information in the highest half of expression/methylation will be compared
## with those in the lowest half, and plots will be generated. The analysis
## will be performed using one CPU core.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to perform the survival analysis
returnValue <- step7TopGenesSurvival(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment
)

## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to perform Kaplan-Meier and Cox regression
## survival analyses on only the top 5 genes and TFs by number of linked
## hypomethylated RE DNA methylation sites, and on all unique
## RE DNA methylation sites linked to those genes. The vital
## status and survival time of patients will be retrieved from a data frame
## with example patient data from the TENET.ExperimentHub package. Gene names
## will be retrieved from the rowRanges of the 'expression'
## SummarizedExperiment object in the example MultiAssayExperiment. In the
## Kaplan-Meier analyses, the patient samples with complete clinical
## information in the highest quartile of expression/methylation will be
## compared with those in the lowest quartile, and plots will not be
## generated. The analysis will be performed using 8 CPU cores.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
```

```
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Load the example clinical data frame from the TENET.ExperimentHub
## package
exampleTENETClinicalDataFrame <-
    TENET.ExperimentHub::exampleTENETClinicalDataFrame()

## Use the example datasets to perform the survival analysis
returnValue <- step7TopGenesSurvival(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    hypermethGplusAnalysis = FALSE,
    topGeneNumber = 5,
    vitalStatusData = exampleTENETClinicalDataFrame$vital_status,
    survivalTimeData = exampleTENETClinicalDataFrame$time,
    highProportion = 0.25,
    lowProportion = 0.25,
    generatePlots = FALSE,
    coreCount = 8
)

## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to perform Kaplan-Meier and Cox regression
## survival analyses on the top 10 genes and TFs by number of linked hyper-
## and hypomethylated RE DNA methylation sites, and on all unique RE DNA
## methylation sites linked to those genes. The vital status and
## survival time of patients will be taken from the "vital_status" and "time"
## columns of the colData of the example MultiAssayExperiment. Gene names
## will be retrieved from the rowRanges of the 'expression'
## SummarizedExperiment object in the example MultiAssayExperiment. In the
## Kaplan-Meier analyses, custom group cutoffs representing quartiles will be
## used, and plots will be generated. The analysis will be performed using
## one CPU core.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Create an example cutoff matrix which will split the samples into
## quartiles and define custom names for the resulting groups
cutoffMatrix <- data.frame(
    "Low" = c(0, (1 / 4), (1 / 2), (3 / 4)),
    "High" = c((1 / 4), (1 / 2), (3 / 4), 1)
)
rownames(cutoffMatrix) <- c(
    "GroupOne",
    "GroupTwo",
    "GroupThree",
    "GroupFour"
)

## Use the example dataset and cutoffMatrix to perform the survival analysis
```

```
returnValue <- step7TopGenesSurvival(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    survivalGroupingCutoffs = cutoffMatrix
)

## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to perform Kaplan-Meier and Cox regression
## survival analyses on the top 10 genes and TFs by number of linked hyper-
## and hypomethylated RE DNA methylation sites, and on all unique RE DNA
## methylation sites linked to those genes. The vital status and
## survival time of patients will be taken from the "vital_status" and "time"
## columns of the colData of the example MultiAssayExperiment. Gene names
## will be retrieved from the rowRanges of the 'expression'
## SummarizedExperiment object in the example MultiAssayExperiment. In the
## Kaplan-Meier analyses, the samples will be divided into 3 groups using
## Jenks natural breaks optimization, and plots will be generated. The
## analysis will be performed using one CPU core.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to perform the survival analysis
returnValue <- step7TopGenesSurvival(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    jenksBreaksGroupCount = 3
)
```

`step7TopGenesTADTables` *Create tables using user-supplied topologically associating domain (TAD) information which identify the TADs containing each RE DNA methylation site linked to the top genes and transcription factors, as well as other genes in the same TAD as potential downstream targets*

Description

This function takes the top genes and transcription factors (TFs) by number of linked RE DNA methylation sites identified by the `step6DNAMethylationSitesPerGeneTabulation` function, up to the number specified by the user, and generates tables for each of the RE DNA methylation sites linked to them in the hyper- and/or hypomethylated G+ analysis quadrants, as selected by the user. These tables note which of the top genes/TFs each RE DNA methylation site is linked to, which TAD each site lies within, and the number and names of genes which lie within the same TAD.

Usage

```
step7TopGenesTADTables(
```

```
  TENETMultiAssayExperiment,
  TADFiles,
  geneAnnotationDataset = NA,
  DNAMethylationArray = NA,
  hypermethGplusAnalysis = TRUE,
  hypomethGplusAnalysis = TRUE,
  topGeneNumber = 10,
  coreCount = 1
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step5OptimizeLinks` and `step6DNAMethylationSitesPerGeneTabulation` functions.

`TADFiles`
: Specify a data frame, matrix, or GRanges object with information on the TAD compartments of interest, organized in a BED-like manner (see https://genome.ucsc.edu/FAQ/FAQformat.html#format1), or a path to a directory that contains *only* one or more BED-like files, which may optionally be compressed (.gz/.bz2/.xz). **Note:** Data frames and matrices must contain 1-indexed coordinates, and BED-like files must contain 0-indexed coordinates.

`geneAnnotationDataset`
: Specify a gene annotation dataset which is used to identify names for genes by their Ensembl IDs. The argument must be either a GRanges object (such as one imported via `rtracklayer::import`) or a path to a GFF3 or GTF file. Both GENCODE and Ensembl annotations are supported. Other annotation datasets may work, but have not been tested. See the "Input data" section of the vignette for information on the required dataset format. Specify NA to use the gene names listed in the "geneName" column of the elementMetadata of the rowRanges of the "expression" SummarizedExperiment object within the TENETMultiAssayExperiment object. Defaults to NA.

`DNAMethylationArray`
: Specify the name of a DNA methylation probe array supported by the sesameData package (see `?sesameData::sesameData_getManifestGRanges`). If an array is specified, RE DNA methylation sites and their locations in that array's manifest are cross-referenced with RE DNA methylation site IDs included in the rownames of the methylation dataset provided in the "methylation" SummarizedExperiment object within the TENETMultiAssayExperiment object, and only those overlapping will be considered for analysis. If set to NA, all RE DNA methylation sites with locations listed in the rowRanges of the "methylation" SummarizedExperiment object are used. Defaults to NA.

`hypermethGplusAnalysis`
: Set to TRUE to create TAD tables for the RE DNA methylation sites linked to the top genes and TFs by most hypermethylated RE DNA methylation sites with G+ links.

hypomethGplusAnalysis
: Set to TRUE to create TAD tables for the RE DNA methylation sites linked to the top genes and TFs by most hypomethylated RE DNA methylation sites with G+ links.

topGeneNumber
: Specify the number of top genes and TFs, based on the most linked RE DNA methylation sites of a given analysis type, for which to generate TAD tables for the RE DNA methylation sites linked to those genes. Defaults to 10.

coreCount
: Argument passed as the mc.cores argument to mcmapply. See ?parallel::mcmapply for more details. Defaults to 1.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named 'step7TopGenesTADTables' in its metadata containing the output of this function. This list contains hypermethGplus and/or hypomethGplus lists, as selected by the user, which contain lists for the top overall genes and top TF genes. These lists contain data frames listing the top genes/TFs each RE DNA methylation site is linked to and, for each TAD file, whether an RE DNA methylation site was found in a TAD in that file, as well as the gene count and identities of other genes found in the same TAD as each RE DNA methylation site.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to do overlapping for all unique RE DNA
## methylation sites linked to the top 10 genes by number of linked hyper-
## and hypomethylated RE DNA methylation sites, using a GRanges object
## containing topologically associating domain (TAD) data from the
## TENET.ExperimentHub package. Gene names and locations, and the locations
## of RE DNA methylation sites, will be retrieved from the rowRanges of the
## 'expression' and 'methylation' SummarizedExperiment objects in the
## example MultiAssayExperiment. The analysis will be performed using one
## CPU core.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Load the example TAD GRanges object from the TENET.ExperimentHub package
exampleTENETTADRegions <- TENET.ExperimentHub::exampleTENETTADRegions()

## Use the example datasets to perform the TAD overlapping
returnValue <- step7TopGenesTADTables(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    TADFiles = exampleTENETTADRegions
)

## This example also uses the example MultiAssayExperiment, but performs
## overlapping for only RE DNA methylation sites linked to the top 5 genes by
## number of linked hypomethylated RE DNA methylation sites. BED-like files
## containing TAD data are retrieved from the directory "TADData". Gene names
```

```
## and locations are retrieved from the rowRanges of the 'expression' and
## 'methylation' SummarizedExperiment objects in the example
## MultiAssayExperiment, and RE DNA methylation sites and their locations are
## retrieved from the HM450 array via the sesameData package. The analysis is
## performed using 8 CPU cores.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to perform the TAD overlapping
returnValue <- step7TopGenesTADTables(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    TADFiles = "TADData",
    DNAMethylationArray = "HM450",
    hypermethGplusAnalysis = FALSE,
    topGeneNumber = 5,
    coreCount = 8
)
```

`step7TopGenesUCSCBedFiles` *Create BED-formatted interact files which can be loaded on the UCSC Genome Browser to display links between top genes and transcription factors and their linked RE DNA methylation sites*

Description

This function takes the top genes and transcription factors (TFs) by number of linked RE DNA methylation sites identified by the `step6DNAMethylationSitesPerGeneTabulation` function in the hyper- and/or hypomethylated G+ analysis quadrants, up to the number specified by the user, and generates BED-formatted interact files (see https://genome.ucsc.edu/goldenPath/help/interact.html) that can be uploaded to the UCSC Genome Browser (https://genome.ucsc.edu) to visualize the links between each of these genes and the RE DNA methylation sites linked to them for the given analysis type.

Usage

```
step7TopGenesUCSCBedFiles(
  TENETMultiAssayExperiment,
  outputDirectory,
  geneAnnotationDataset = NA,
  DNAMethylationArray = NA,
  hypermethGplusAnalysis = TRUE,
  hypomethGplusAnalysis = TRUE,
  topGeneNumber = 10
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step5OptimizeLinks` and `step6DNAMethylationSitesPerGeneTabulation` functions.

`outputDirectory`
: Specify the path to the output directory in which to save the `.inter.bed` files created by this function. It will be created if necessary.

`geneAnnotationDataset`
: Specify a gene annotation dataset which is used to identify names for genes by their Ensembl IDs. The argument must be either a GRanges object (such as one imported via `rtracklayer::import`) or a path to a GFF3 or GTF file. Both GENCODE and Ensembl annotations are supported. Other annotation datasets may work, but have not been tested. See the "Input data" section of the vignette for information on the required dataset format. Specify NA to use the gene names listed in the "geneName" column of the elementMetadata of the rowRanges of the "expression" SummarizedExperiment object within the TENETMultiAssayExperiment object. Defaults to NA.

`DNAMethylationArray`
: Specify the name of a DNA methylation probe array supported by the sesameData package (see `?sesameData::sesameData_getManifestGRanges`). If an array is specified, RE DNA methylation sites and their locations in that array's manifest are cross-referenced with RE DNA methylation site IDs included in the rownames of the methylation dataset provided in the "methylation" SummarizedExperiment object within the TENETMultiAssayExperiment object, and only those overlapping will be considered for analysis. If set to NA, all RE DNA methylation sites with locations listed in the rowRanges of the "methylation" SummarizedExperiment object are used. Defaults to NA.

`hypermethGplusAnalysis`
: Set to TRUE to create interact files showing links between the top genes and TFs by most RE hypermethylated RE DNA methylation sites with G+ links and their linked RE DNA methylation sites. Defaults to TRUE.

`hypomethGplusAnalysis`
: Set to TRUE to create interact files showing links between the top genes and TFs by most hypomethylated RE DNA methylation sites with G+ links and their linked RE DNA methylation sites. Defaults to TRUE.

`topGeneNumber`
: Specify the number of top genes and TFs, based on the most linked RE DNA methylation sites of a given analysis type, for which to generate interact files showing the links between those genes and each of their linked RE DNA methylation sites. Defaults to 10.

Value

Outputs BED-formatted interact files to upload to the UCSC Genome Browser to the specified output directory. These files display the interactions between the top genes/TFs and their linked RE DNA methylation sites for the given analysis types. Returns a list of lists named after each selected

analysis type, each containing the file paths to the created .inter.bed files for top genes and top TFs for that analysis type.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to create UCSC Genome Browser interact files
## for the top 10 genes and TFs by number of linked hyper- and hypomethylated
## RE DNA methylation sites. The interact files for the top genes and TFs
## will be saved in the user's working directory. Gene names and locations,
## and the locations of RE DNA methylation sites, will be retrieved from the
## rowRanges of the 'expression' and 'methylation' SummarizedExperiment
## objects in the example MultiAssayExperiment.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to create the UCSC Genome Browser interact files
filePaths <- step7TopGenesUCSCBedFiles(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    outputDirectory = "."
)

## Get the path to the .inter.bed file for the top TFs by number of
## hypomethylated G+ RE DNA methylation sites
filePaths$hypoGplus$topTFs

## This example is similar, but creates UCSC Genome Browser interact files
## for only the top 5 genes and TFs by number of linked hypomethylated RE DNA
## methylation sites, and RE DNA methylation site IDs and locations are
## retrieved from the HM450 array via the sesameData package.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example dataset to create the UCSC Genome Browser interact files
filePaths <- step7TopGenesUCSCBedFiles(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    outputDirectory = ".",
    DNAMethylationArray = "HM450",
    hypermethGplusAnalysis = FALSE,
    topGeneNumber = 5
)

## Get the path to the .inter.bed file for the top TFs by number of
## hypomethylated G+ RE DNA methylation sites.
## Note: Since we performed analyses only using TFs in the step 3 function,
## the top genes are all TFs, so topTFs will be NA and topGenes must be used
## instead.
```

```
filePaths$hypoGplus$topGenes
```

`step7TopGenesUserPeakOverlap`

Identify if RE DNA methylation sites linked to top genes and transcription factors are located within a specific distance of specified genomic regions

---

Description

This function takes the top genes and transcription factors (TFs) by number of linked RE DNA methylation sites identified by the `step6DNAMethylationSitesPerGeneTabulation` function, up to the number specified by the user, and identifies if the RE DNA methylation sites linked to those genes/TFs from the hyper- and/or hypomethylated G+ analysis quadrants are found in the vicinity of genomic regions (peaks) of interest, supplied by the user in the form of .bed, .narrowPeak, .broadPeak, and/or gappedPeak files, directories containing these files, data frames, and/or GRanges objects.

Usage

```
step7TopGenesUserPeakOverlap(
  TENETMultiAssayExperiment,
  peakData,
  geneAnnotationDataset = NA,
  DNAMethylationArray = NA,
  hypermethGplusAnalysis = TRUE,
  hypomethGplusAnalysis = TRUE,
  topGeneNumber = 10,
  distanceFromREDNAMethylationSites = 100,
  coreCount = 1
)
```

Arguments

`TENETMultiAssayExperiment`
: Specify a MultiAssayExperiment object containing expression and methylation SummarizedExperiment objects, such as one created by the TCGADownloader function. The object's metadata must contain the results from the `step5OptimizeLinks` and `step6DNAMethylationSitesPerGeneTabulation` functions.

`peakData`
: Specify a data frame, matrix, or GRanges object with genomic regions (peaks) of interest, organized in a BED-like manner (see https://genome.ucsc.edu/FAQ/FAQformat.html#format1), a path to a .bed, .narrowPeak, .broadPeak, and/or .gappedPeak file with peaks of interest, a path to a directory containing one or more of these file types, or a named list of any of these types of input.

Peak names are taken from the fourth column of the input if it exists, or, if the input is a GRanges object, the names of the ranges. Additional columns can be included, but are not used by this function. If no names are present, they are generated from peak coordinates and take the form <chromosome>_<start>_<end>[.<optionalDuplicate
Input files may optionally be compressed (.gz/.bz2/.xz).

geneAnnotationDataset

Specify a gene annotation dataset which is used to identify names for genes by their Ensembl IDs. The argument must be either a GRanges object (such as one imported via rtracklayer::import) or a path to a GFF3 or GTF file. Both GENCODE and Ensembl annotations are supported. Other annotation datasets may work, but have not been tested. See the "Input data" section of the vignette for information on the required dataset format. Specify NA to use the gene names listed in the "geneName" column of the elementMetadata of the rowRanges of the "expression" SummarizedExperiment object within the TENETMultiAssayExperiment object. Defaults to NA.

DNAMethylationArray

Specify the name of a DNA methylation probe array supported by the sesameData package (see ?sesameData::sesameData_getManifestGRanges). If an array is specified, RE DNA methylation sites and their locations in that array's manifest are cross-referenced with RE DNA methylation site IDs included in the rownames of the methylation dataset provided in the "methylation" SummarizedExperiment object within the TENETMultiAssayExperiment object, and only those overlapping will be considered for analysis. If set to NA, all RE DNA methylation sites with locations listed in the rowRanges of the "methylation" SummarizedExperiment object are used. Defaults to NA.

hypermethGplusAnalysis

Set to TRUE to create data frames with the peak overlap information for the unique hypermethylated RE DNA methylation sites linked to the top genes and TFs by most hypermethylated RE DNA methylation sites with G+ links. Defaults to TRUE.

hypomethGplusAnalysis

Set to TRUE to create data frames with the peak overlap information for the unique hypomethylated RE DNA methylation sites linked to the top genes and TFs by most hypomethylated RE DNA methylation sites with G+ links. Defaults to TRUE.

topGeneNumber Specify the number of top genes and TFs, based on the most linked RE DNA methylation sites of a given analysis type, for which to generate data showing overlap with the specified peak datasets for the RE DNA methylation sites linked to those genes. Defaults to 10.

distanceFromREDNAMethylationSites

Specify the distance from the linked RE DNA methylation sites within which an RE DNA methylation site will be considered to overlap a peak. Must be a nonnegative integer. Defaults to 100.

coreCount Argument passed as the mc.cores argument to mclapply. See ?parallel::mclapply for more details. Defaults to 1.

Value

Returns the MultiAssayExperiment object given as the TENETMultiAssayExperiment argument with an additional list named 'step7TopGenesUserPeakOverlap' in its metadata containing the output of this function. This list contains `hypermethGplus` and/or `hypomethGplus` lists, as selected by the user, which contain lists for the top overall genes and top TF genes. Each of these lists contains two elements. The first, `peakDatasetOverlapInfo`, is a list of data frames named after the peak datasets (without file extensions). If a single R object was provided as input, the list will contain a single element named 'peakData'. Each data frame contains peak names in the column names and RE DNA methylation site IDs in the row names. The Boolean values indicate whether each RE DNA methylation site overlaps with each peak. The second, `linkedDNAMethylationSiteInfo`, is a data frame containing a row for each of the unique RE DNA methylation sites linked to the top genes/TFs for the specified analysis types. The columns note the location of the RE DNA methylation site, the specified search window for the site, and whether the site is linked to each of the top genes/TFs.

Examples

```
## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to overlap example peaks with all unique RE
## DNA methylation sites linked to the top 10 genes by number of linked
## hyper- and hypomethylated RE DNA methylation sites, using a GRanges object
## containing the genomic coordinates of peaks of interest. Gene names and
## the locations of RE DNA methylation sites will be retrieved from the
## rowRanges of the 'expression' and 'methylation' SummarizedExperiment
## objects in the example MultiAssayExperiment. A window of 100 base pairs
## will be used to identify if the RE DNA methylation sites lie within the
## vicinity of peaks. The analysis will be performed using one CPU core.

## Load the example TENET MultiAssayExperiment object
## from the TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Load the example peak GRanges object from the TENET.ExperimentHub package
exampleTENETPeakRegions <- TENET.ExperimentHub::exampleTENETPeakRegions()

## Use the example datasets to perform the peak overlapping
returnValue <- step7TopGenesUserPeakOverlap(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    peakData = exampleTENETPeakRegions
)

## This example uses the example MultiAssayExperiment provided in the
## TENET.ExperimentHub package to overlap specified peaks with all unique RE
## DNA methylation sites linked to only the top 5 genes by number of linked
## hypomethylated RE DNA methylation sites. The genomic coordinates of peaks
## of interest will be loaded from BED-like files located in the user's R
## working directory. Gene names will be retrieved from the rowRanges of the
## 'expression' SummarizedExperiment object in the example
## MultiAssayExperiment, and RE DNA methylation sites and their locations
## will be retrieved from the HM450 array via the sesameData package. A
```

```
## window of 500 base pairs will be used to identify if the RE DNA
## methylation sites lie within the vicinity of peaks. The analysis will be
## performed using 8 CPU cores.

## Load the example TENET MultiAssayExperiment object from the
## TENET.ExperimentHub package
exampleTENETMultiAssayExperiment <-
    TENET.ExperimentHub::exampleTENETMultiAssayExperiment()

## Use the example datasets to perform the peak overlapping
returnValue <- step7TopGenesUserPeakOverlap(
    TENETMultiAssayExperiment = exampleTENETMultiAssayExperiment,
    peakData = ".",
    DNAMethylationArray = "HM450",
    hypermethGplusAnalysis = FALSE,
    topGeneNumber = 5,
    distanceFromREDNAMethylationSites = 500,
    coreCount = 8
)
```

`TCGADownloader` *Download TCGA gene expression, DNA methylation, and clinical datasets and compile them into a MultiAssayExperiment object*

Description

This function downloads and compiles TCGA gene expression and DNA methylation datasets, as well as clinical data primarily intended for use with the TENET package. This simplifies the TCGAbiolinks download functions, identifies samples with matching gene expression and DNA methylation data, and can also remove duplicate tumor samples taken from the same patient donor. Data are compiled into a MultiAssayExperiment object, which is returned and optionally saved in an `.rda` file at the path specified by the `outputFile` argument.

Usage

```
TCGADownloader(
  rawDataDownloadDirectory,
  GDCDownloadMethod = "api",
  filesPerChunk = 10,
  TCGAStudyAbbreviation,
  RNASeqWorkflow,
  RNASeqLog2Normalization = TRUE,
  removeDupTumor = TRUE,
  matchingExpAndMetSamples = TRUE,
  clinicalSurvivalData = "combined",
  outputFile = NA
)
```

Arguments

`rawDataDownloadDirectory`
: Specify the path to the directory where TCGAbiolinks should download data. **Note:** The downloaded files can be very large.

`GDCDownloadMethod`
: The method to use when downloading data from the Genomic Data Commons (GDC). Passed as the `method` argument to TCGAbiolinks' `GDCdownload` function. The available options are "api" and "client"; the default is "api". The "api" method works on all operating systems, but it does not retry the download of incomplete or corrupted files, so `TCGADownloader` must be manually rerun in this case. The "client" method is more reliable, but it requires Windows, macOS (Apple Silicon only), or Ubuntu (64-bit x86 only), or manual installation of the GDC Data Transfer Tool Client (which must be in the command search path).

`filesPerChunk`
: The number of data files to download at once when using the "api" download method. Passed as the `files.per.chunk` argument to TCGAbiolinks' `GDCdownload` function. Lower values may improve download reliability, but higher values may increase download speed. Defaults to 10.

`TCGAStudyAbbreviation`
: Specify the four-letter abbreviation of a TCGA dataset for which to download data. See https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations for more information and a complete list of options.

`RNASeqWorkflow`
: Select the type of RNA-seq data to download. For TENET purposes, choose either "STAR - FPKM", "STAR - FPKM-UQ", "STAR - FPKM-UQ - old formula", or "STAR - TPM". "STAR - Counts" may also be used but is not recommended for TENET analyses. See https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/ for the meaning of these options. "STAR - FPKM-UQ - old formula" is specific to TENET; it uses "STAR - FPKM-UQ", but multiplies the FPKM-UQ values by 19,029 (the number of human protein coding genes on autosomes), resulting in values similar to those TCGA used prior to Data Release 37.0 on March 29, 2023. This allows the comparison of TCGA FPKM-UQ datasets downloaded before and after that date.

`RNASeqLog2Normalization`
: Set to TRUE to perform log2 normalization of RNA-seq expression values. Defaults to TRUE.

`removeDupTumor`
: Set to TRUE to remove duplicate tumor samples taken from the same subject, leaving only one sample per subject in alphanumeric order. **Note:** To properly create a dataset for use with TENET, both the `removeDupTumor` and `matchingExpAndMetSamples` arguments must be set to TRUE. Defaults to TRUE.

`matchingExpAndMetSamples`
: If set to TRUE, only data for patients with at least one methylation and expression sample will be kept. If set to FALSE, all samples will be kept. **Note:** To properly create a dataset for use with TENET, both the removeDupTumor and matchingExpAndMetSamples arguments must be set to TRUE. Defaults to TRUE.

`clinicalSurvivalData`
: Select how patient vital status and survival time data should be extracted from the TCGA data. Specify "bcrBiotabPatient" to use survival data from only the 'patient' dataset in the BCR Biotab files downloaded using TCGAbiolinks, or "combined" to use survival data from the 'patient' and 'follow_up' datasets in the BCR Biotab files, as well as the BCR XML files. Data from the same patient in each of the datasets are combined, and the most recent entry (highest patient survival time) for each patient is kept. For both options, the 'days_to_last_followup' and 'days_to_death' variables are collapsed into a single time variable, which is combined with the other clinical data in the 'patient' BCR Biotab data. See https://bioconductor.org/packages/devel/bioc/vignettes/TCGAbiolinks/inst/doc/clinical.html for more information on how TCGAbiolinks prepares clinical datasets. Defaults to "combined".

`outputFile`
: Specify the path to an `.rda` file in which to save the created MultiAssayExperiment object. If set to NA, the object is only returned. Defaults to NA.

Value

Returns a MultiAssayExperiment object containing SummarizedExperiment objects with expression and methylation data, as well as clinical data in its colData.

Examples

```
## This example downloads a TCGA LUAD dataset with log2-normalized
## FPKM-UQ expression values from tumor and adjacent normal tissue samples
## with matching expression and methylation data, keeping only one tumor
## sample from each patient. Survival data will be combined from three
## clinical datasets downloaded by TCGAbiolinks. Raw data files will be saved
## to the R working directory, and the processed dataset will only be
## returned as a variable.
TCGADataset <- TCGADownloader(
    rawDataDownloadDirectory = ".",
    TCGAStudyAbbreviation = "LUAD",
    RNASeqWorkflow = "STAR - FPKM-UQ"
)

## This example downloads a TCGA BRCA dataset with FPKM expression values
## with no normalization and does not remove duplicate samples. Survival
## data are derived from only the patient BCR Biotab file downloaded by
## TCGAbiolinks. Both raw data files and an .rda file containing the data
## as a MultiAssayExperiment object will be saved to the R working directory.
## Note: The resulting object will *not* work for a TENET analysis due to the
## lack of sample matching and duplicate tumor sample removal.
TCGADownloader(
    rawDataDownloadDirectory = ".",
    TCGAStudyAbbreviation = "BRCA",
    RNASeqWorkflow = "STAR - FPKM",
    RNASeqLog2Normalization = FALSE,
    removeDupTumor = FALSE,
    matchingExpAndMetSamples = FALSE,
    clinicalSurvivalData = "bcrBiotabPatient",
```

```
    outputFile = "BRCAMultiAssayExperimentObject.rda"
)
```

TENET *TENET (Tracing regulatory Element Networks using Epigenetic Traits)*

Description

TENET identifies key transcription factors and regulatory elements linked to a specific cell type by finding significantly correlated differences in gene expression and regulatory element methylation between case and control input datasets, and identifying the top genes by number of significant RE DNA methylation site links. It also includes many additional tools to aid in the visualization and analysis of the results, including plots displaying and comparing methylation and expression data and RE DNA methylation site link counts, survival analysis, TF motif searching in the vicinity of linked and specified RE DNA methylation sites and within specified regions, custom TAD and peak overlap analysis, and UCSC Genome Browser track file generation. A utility function is also provided to download methylation, expression, and patient survival data from The Cancer Genome Atlas (TCGA) for use in TENET or other analyses.

Author(s)

Maintainer: Rhie Lab at the University of Southern California <rhielab@gmail.com>

Authors:

- Daniel Mullen (ORCID)
- Zexun Wu (ORCID)
- Ethan Nelson-Moore (ORCID)
- Suhn Rhie (ORCID)

See Also

Useful links:

- https://github.com/rhielab/TENET
- Report bugs at https://github.com/rhielab/TENET/issues

`TENETCacheAllData` *Cache all online datasets required by TENET examples and optional features*

Description

This function locally caches all online TENET and SeSAMe datasets required by TENET examples and optional features (TENET.ExperimentHub objects used in examples, TENET.AnnotationHub datasets used in step 1, and SeSAMe datasets loaded via the `DNAMethylationArray` argument). The main purpose of this function is to enable the use of TENET in an environment without internet access, such as the compute nodes of an HPC cluster. In this case, you must run `TENETCacheAllData()` once while connected to the internet before using TENET examples or these optional features.

Usage

```
TENETCacheAllData()
```

Value

Returns NULL.

Examples

```
TENETCacheAllData()
```

Index

∗ **datasets**
  humanTranscriptionFactorDb, 6
  humanTranscriptionFactorList, 8
∗ **internal**
  TENET, 64

easyTENET, 3

humanTranscriptionFactorDb, 6
humanTranscriptionFactorList, 8

step1MakeExternalDatasets, 8
step2GetDifferentiallyMethylatedSites, 11
step3GetAnalysisZScores, 15
step4SelectMostSignificantLinksPerDNAMethylationSite, 17
step5OptimizeLinks, 19
step6DNAMethylationSitesPerGeneTabulation, 21
step7ExpressionVsDNAMethylationScatterplots, 23
step7LinkedDNAMethylationSiteCountHistograms, 27
step7LinkedDNAMethylationSitesMotifSearching, 29
step7SelectedDNAMethylationSitesCaseVsControlBoxplots, 33
step7StatesForLinks, 35
step7TopGenesCaseVsControlExpressionBoxplots, 36
step7TopGenesCircosPlots, 39
step7TopGenesDNAMethylationHeatmaps, 41
step7TopGenesExpressionCorrelationHeatmaps, 43
step7TopGenesOverlappingLinkedDNAMethylationSitesHeatmaps, 45
step7TopGenesSurvival, 47
step7TopGenesTADTables, 52
step7TopGenesUCSCBedFiles, 55
step7TopGenesUserPeakOverlap, 58

TCGADownloader, 61
TENET, 64
TENET-package *(*TENET*)*, 64
TENETCacheAllData, 65