Package ‘TCGAWorkflowData’

May 14, 2024

Title Data for TCGA Workflow

Version 1.28.0

Author Tiago Chedraoui Silva <tiagochst@gmail.com>

Maintainer Tiago Chedraoui Silva <tiagochst@gmail.com>

Description This experimental data package contains 11 data sets necessary to follow the "TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages".

Depends R (>= 3.5.0)

Imports SummarizedExperiment

License GPL-3

VignetteBuilder knitr

biocViews ExperimentData, Homo_sapiens_Data, MicroarrayData, CancerData

NeedsCompilation no

URL https://f1000research.com/articles/5-1542/v2

BugReports https://github.com/BioinformaticsFMRP/TCGAWorkflow/issues

RoxygenNote 7.2.3

Suggests knitr, rmarkdown, pander, testthat, BiocStyle

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

git_branch RELEASE_3_19

git_last_commit ce36f3e

git_last_commit_date 2024-04-30

Repository Bioconductor 3.19

Date/Publication 2024-05-14
### exp

A gene expression matrix for 10 GBM and 10 LGG samples prepared for the creation of an ELMER object.

#### Description

A gene expression matrix for 10 GBM and 10 LGG samples prepared for the creation of an ELMER object.

#### Format

A matrix with 21022 rows and 20 samples

#### Examples

```r
data("elmerExample")
```

### exp_gbm

A gene expression matrix with 20 GBM samples

#### Description

A gene expression matrix with 20 GBM samples

#### Format

Gene expression: A SummarizedExperiment object with 21022 rows and 20 columns
exp_lgg

Examples

data("TCGA_GBM_Transcriptome_20_samples")

desc exp_lgg

A gene expression matrix with 20 LGG samples

Description

A gene expression matrix with 20 LGG samples

Format

Gene expression: A SummarizedExperiment object with 21022 rows and 20 columns

Examples

data("TCGA_LGG_Transcriptome_20_samples")

desc gbm.samples

Identifiers for the 10 GBM samples in the ELMER objects

Description

Identifiers for the 10 GBM samples in the ELMER objects

Format

A vector of 10 barcodes

Examples

data("elmerExample")

desc genes

A data frame object with gene information (hg19)

Description

A data frame object with gene information (hg19)

Format

A dataframe object

Examples

data("genes_GR")
genes_GR

A GRanges object with gene information (hg19)

Description

A GRanges object with gene information (hg19)

Format

A GRanges object

Examples

data("genes_GR")

gistic_allbygene

A subset of GBM GISTIC2 results, which is used to identify genes targeted by somatic copy-number alterations (SCNAs) From: GDAC firehose, downloaded with RTCGAtoolbox

Description

A subset of GBM GISTIC2 results, which is used to identify genes targeted by somatic copy-number alterations (SCNAs) From: GDAC firehose, downloaded with RTCGAtoolbox

Format

A matrix with 24776 rows and 580 columns

Examples

data("gbm_gistic")
gistic_thresholedbygene

A subset of GBM GISTIC2 results, which is used to identify genes targeted by somatic copy-number alterations (SCNAs) From: GDAC firehose, downloaded with RTCGAtoolbox

Description

A subset of GBM GISTIC2 results, which is used to identify genes targeted by somatic copy-number alterations (SCNAs) From: GDAC firehose, downloaded with RTCGAtoolbox

Format

A matrix with 24776 rows and 580 columns

Examples

data("gbm_gistic")

histone.marks

Histone marks specific for brain tissue from the Roadmap database.

Description

Histone marks specific for brain tissue from the Roadmap database.

Format

A matrix with 72102 rows and 114 columns

Examples

data("histoneMarks")
<table>
<thead>
<tr>
<th>lgg.samples</th>
<th>Identifiers for the 10 LGG samples in the ELMER objects</th>
</tr>
</thead>
</table>

**Description**

Identifiers for the 10 LGG samples in the ELMER objects

**Format**

A vector of 10 barcodes

**Examples**

data("elmerExample")

<table>
<thead>
<tr>
<th>maf</th>
<th>Merged LGG and GBM GDC MAF files from GDC workflow: Aliquot Ensemble Somatic Variant Merging and Masking</th>
</tr>
</thead>
</table>

**Description**

Merged LGG and GBM GDC MAF files from GDC workflow: Aliquot Ensemble Somatic Variant Merging and Masking

**Format**

A matrix with 87957 rows and 141 columns

**Examples**

data("maf_lgg_gbm")

<table>
<thead>
<tr>
<th>met</th>
<th>A SummarizedExperiment containing TCGA data: DNA methylation platform 450K chromosome 9 for 10 LGG samples and 10 GBM samples</th>
</tr>
</thead>
</table>

**Description**

A SummarizedExperiment containing TCGA data: DNA methylation platform 450K chromosome 9 for 10 LGG samples and 10 GBM samples

**Format**

A SumarizedExperiment with 9861 rows and 20 samples
**TCGAWorkflowData**

**Examples**

```r
data("elmerExample")
```

---

**Description**

This experimental data package has the data necessary to follow the TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages. It contains the following files:

- **met20SamplesGBMLGGchr9**: DNA methylation matrix from Infinium HumanMethylation450 platform for 10 LGG (Lower grade glioma) and 10 GBM (Glioblastoma multiforme). It has only probes in chromosome 9 in order to make the example of the workflow faster.

- **elmerExample**: Contains a DNA methylation matrix (only probes in chromosome 9) and a gene expression matrix for 10 LGG (Lower grade glioma) and 10 GBM (Glioblastoma multiforme) in the required format for to execute the R/Bioconductor ELMER package analysis and a vector identifying which sample belongs to each tumor type.

- **biogrid**: biogrid information

- **maf_lgg_gbm**: Mutation annotation files for LGG (Lower grade glioma) and GBM (Glioblastoma multiforme) samples merged into a single matrix. The GDC Somatic Mutation Calling Workflow mutect2 was used to create this MAF files.

- **histoneMarks**: histone marks specific for brain tissue using from Roadmap database.

- **genes_GR**: A GRanges Object and a dataframe with gene information (hg19) downloaded from ENSEMBLE database using biomart via TCGAbiolinks

- **TCGA_GBM_Transcriptome_20_samples**: a matrix with raw expression signal for expression of a gene for 20 GBM (Glioblastoma multiforme) samples

- **TCGA_LGG_Transcriptome_20_samples**: a matrix with raw expression signal for expression of a gene for 20 LGG (low grade glioma) samples

For more information how to create these objects please read the vignette of this package with the following command: `browseVignettes("TCGAWorkflowData")`

---

**Examples**

```r
data("elmerExample")
data("TCGA_LGG_Transcriptome_20_samples")
data("TCGA_GBM_Transcriptome_20_samples")
data("histoneMarks")
data("biogrid")
data("genes_GR")
data("maf_lgg_gbm")
```
Description

Biogrid information

Format

Two matrices with 24776 rows and 580 columns

Examples

data("biogrid")
Index

* internal
  exp, 2
  exp_gbm, 2
  exp_lgg, 3
  gbm.samples, 3
  genes, 3
  genes_GR, 4
  gistic_allbygene, 4
  gistic_threholedbygene, 5
  histone.marks, 5
  lgg.samples, 6
  maf, 6
  met, 6
  tmp.biogrid, 8

* utilities
  TCGAWorkflowData, 7
  exp, 2
  exp_gbm, 2
  exp_lgg, 3
  gbm.samples, 3
  genes, 3
  genes_GR, 4
  gistic_allbygene, 4
  gistic_threholedbygene, 5
  histone.marks, 5
  lgg.samples, 6
  maf, 6
  met, 6
  TCGAWorkflowData, 7
  tmp.biogrid, 8