

# Package ‘TCGAWorkflowData’

April 11, 2023

**Title** Data for TCGA Workflow

**Version** 1.22.0

**Author** Tiago Chedraoui Silva <tiagochst@usp.br>

**Maintainer** Tiago Chedraoui Silva <tiagochst@usp.br>

**Description** This experimental data package contains 11 data sets necessary to follow the  
``TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor pack-  
ages''.

**Depends** R (>= 3.4.0)

**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** 6.0.1

**Suggests** knitr, rmarkdown, pander, testthat

**git\_url** <https://git.bioconductor.org/packages/TCGAWorkflowData>

**git\_branch** RELEASE\_3\_16

**git\_last\_commit** f44df94

**git\_last\_commit\_date** 2022-11-01

**Date/Publication** 2023-04-11

## R topics documented:

TCGAWorkflowData . . . . .	2
<b>Index</b>	<b>4</b>

## 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:

- `GBMnocnvhg19`: segmented CNV from SNP array (Affymetrix Genome-Wide Human SNP Array 6.0) for 20 Glioblastoma multiforme (GBM) samples.
- `GBMIllumina_HiSeq`: a matrix with raw expression signal for expression of a gene for 20 GBM (Glioblastoma multiforme) samples
- `LGGIllumina_HiSeq`: a matrix with raw expression signal for expression of a gene for 20 LGG (Lower grade glioma) samples
- `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.
- `GBMGistic`: a subset of GBM GISTIC results which is used to identify genes targeted by somatic copy-number alterations (SCNAs)
- `biogrid`: biogrid information
- `mafMutect2LGGGBM`: 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.
- `markersMatrix`: Probes meta file from broadinstitute website for Copy Number Variation Analysis (CNV) analysis
- `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

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

## Examples

```
data("elmerExample")
data("GBMnocnvhg19")
data("GBMIllumina_HiSeq")
data("LGGIllumina_HiSeq")
data("met20SamplesGBMLGGchr9")
data("mafMutect2LGGGBM")
```

```
data("markersMatrix")  
data("biogrid")  
data("GBMGistic")  
data("histoneMarks")  
data("genes_GR")
```

# Index

\* **utilities**

TCGWorkflowData, [2](#)

TCGWorkflowData, [2](#)

TCGWorkflowData-package  
(TCGWorkflowData), [2](#)