Package 'HarmonizedTCGAData'

November 13, 2025

Type Package

Title Processed Harmonized TCGA Data of Five Selected Cancer Types

Version 1.32.0

Author Tianle Ma

Maintainer Tianle Ma <tianlema@buffalo.edu>

Description This package contains the processed harmonized TCGA data of five cancer types used in ``Tianle Ma and Aidong Zhang, Integrate Multi-omic Data Using Affinity Network Fusion (ANF) for Cancer Patient Clustering".

License GPL-3

Imports ExperimentHub

Depends R (>= 3.4.0)

VignetteBuilder knitr

Suggests AnnotationHub, knitr, rmarkdown, igraph, survival, ANF, stats, testthat

biocViews CancerData, ReproducibleResearch

RoxygenNote 6.0.1

LazyData true

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

git_branch RELEASE_3_22

git_last_commit 36dd1c5

git_last_commit_date 2025-10-29

Repository Bioconductor 3.22

Date/Publication 2025-11-13

Contents

HarmonizedTCGAData *HarmonizedTCGAData.*

Description

HarmonizedTCGAData.

project_ids *project_ids*

Description

A named character vector: mapping case_id (i.e., patient ID) to the TCGA project ID they belong to

Usage

```
project_ids
```

Format

An object of class `character` of length 14551.

Examples

```
library(ExperimentHub)
eh <- ExperimentHub()
myfiles <- query(eh, "HarmonizedTCGAData")
project_ids <- myfiles[[2]]
# project_ids <- myfiles[['EH1015']]
head(project_ids)
```

surv.plot *surv.plot*

Description

Patient survival information (overall survival plot data) were downloaded from https://portal.gdc.cancer.gov/exploration?s
For detailed information: see section "Survival analysis" in https://docs.gdc.cancer.gov/Data_Portal/PDF/Data_Portal_U

Usage

```
surv.plot
```

Format

A data frame with four variables: `survivalEstimate`, `id`, `censored`, and `time`

Examples

```
library(ExperimentHub)
eh <- ExperimentHub()
myfiles <- query(eh, "HarmonizedTCGAData")
surv.plot <- myfiles[[3]]
# surv.plot <- myfiles[['EH1016']]
head(surv.plot)
```

`Wall` *Wall*

Description

Wall contains a list of precomputed affinity (similarity) matrices of 2582 patients. These matrices were derived from 10382 gene expression, miRNA expression and DNA methylation data files downloaded from GDC data portal The file UUIDs can be found in inst/extdata/fileUUIDs.csv Using these file UUIDs, users can download the original data from https://portal.gdc.cancer.gov/repository 'Wall' is a complex list and contains lists inside list. Precisely, Wall is a list (five cancer types) of list (six feature normalization types: raw.all, raw.sel, log.all, log.sel, vst.sel, normalized) of list (three feature spaces or views: fpkm, mirna, and methy450) of matrices. (So Wall contains 90 matrices in total) The rownames of each matrix is the case_id (i.e., patient id), and the column names of each matrix is the aliquot IDs (i.e., TCGA barcode, which contains the case_id as prefix).

Usage

```
Wall
```

Format

An object of class `list` of length 5.

Examples

```
library(ExperimentHub)
eh <- ExperimentHub()
myfiles <- query(eh, "HarmonizedTCGAData")
Wall <- myfiles[[1]]
# Wall <- myfiles[['EH1014']]
names(Wall)
names(Wall[[1]])
names(Wall[[1]][[1]])
dim(Wall[[1]][[1]][[1]])
```

Index