

Package: NGCHMDemoData (via r-universe)

March 18, 2025

Title Demo Data for the NGCHM R Package

Version 1.0.0

Description Package of demo data for NGCHM vignettes.

Depends R (>= 3.5.0)

License GPL-3

URL <https://github.com/MD-Anderson-Bioinformatics/NGCHMDemoData>,
<https://md-anderson-bioinformatics.r-universe.dev/NGCHMDemoData>

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

Data TCGA.BRCA.Expression.csv, TCGA.BRCA.TP53Mutation.csv

Repository <https://md-anderson-bioinformatics.r-universe.dev>

RemoteUrl <https://github.com/MD-Anderson-Bioinformatics/NGCHMDemoData>

RemoteRef HEAD

RemoteSha a224a5261efbe89c7d1884c1a2fbf6e0d6bc7efa

Contents

NGCHMDemoData	2
TCGA.BRCA.Demo	3
TCGA.BRCA.ExpressionData	3
TCGA.BRCA.TP53MutationData	4
TCGA.GBM.Demo	4
TCGA.GBM.EXPR	5
TCGA.GBM.ExpressionData	5
TCGA.GBM.TP53MutationData	6

NGCHMDemoData

NGCHMDemoData

Description

This package provides several relatively large datasets that can be used to demonstrate the capabilities of the Next-Generation Clustered Heat Map (NG-CHM) system.

Details

The included data is a small subset of data from The Cancer Genome Atlas (TCGA) project. There are five data files from three groups of cancer samples:

- 200 breast cancer (BRCA) samples.
- 169 glioblastoma (GBM) samples.
- 547 additional glioblastoma (GBM) samples.

The two Glioblastoma groups were characterized using different technologies (RNASeq and microarrays, respectively).

Note: the NG-CHM system can work with data from any domain, not just biological data.

Note: the included data has been preprocessed, subsetted, and manipulated in multiple, undocumented ways. It should only be used for evaluating and demonstrating the NG-CHM system and not for deriving any scientific conclusions.

The different data sets overlap with each other in several ways that are documented in the data sets concerned. These overlaps can be easily used to generate NG-CHMs that integrate multiple data sets in a variety of ways to further demonstrate the capabilities of NG-CHMs.

Installation

This package can be installed from MD Anderson Bioinformatics R-universe repository:

```
install.packages("NGCHMDemoData",  
repos = c("https://md-anderson-bioinformatics.r-universe.dev", "https://cran.r-project.org"))
```

Author(s)

Maintainer: Mary A Rohrdanz <marohrdanz@mdanderson.org>

Authors:

- Bradley M Broom <bmbroom@mdanderson.org> ([ORCID](#))

Other contributors:

- Chris Wakefield <cwakefield@mdanderson.org> [contributor]
- James Melott <jmmelott@mdanderson.org> [contributor]
- MD Anderson Cancer Center [copyright holder]

See Also

TCGA.BRCA.Demo, TCGA.GBM.Demo, TCGA.GBM.EXPR

TCGA.BRCA.Demo	<i>A subset of the breast cancer (BRCA) data from TCGA.</i>
----------------	---

Description

This dataset is loaded automatically when the package is loaded. It consists of two related parts:

- A matrix of gene expression data.
- A vector containing the TP53 mutation status of each sample in the matrix.

See Also

TCGA.BRCA.ExpressionData, TCGA.BRCA.TP53MutationData

TCGA.BRCA.ExpressionData	<i>A subset of the breast cancer (BRCA) expression data from TCGA</i>
--------------------------	---

Description

A subset of the breast cancer (BRCA) expression data from TCGA

Format

A numeric data matrix with 3437 rows and 200 columns.

- Row labels are gene symbols (e.g. TSPAN6). The NG-CHM label type is `bio.gene.hugo`.
- Column labels are TCGA barcodes up to the sample/vial field (16 characters total, e.g. TCGA-AO-A0JJ-01A). The NG-CHM label type is `bio.tcga.barcode.sample.vial`.
- Data has been log-transformed (min 1, max 21.75322)

See Also

TCGA.BRCA.Demo, TCGA.BRCA.TP53MutationData

TCGA.BRCA.TP53MutationData

TP53 mutation data for TCGA breast cancer (BRCA) samples

Description

TP53 mutation data for TCGA breast cancer (BRCA) samples

Format

A length 200 character vector.

- Each element of the vector is either "WT" or "MUT".
- Element names are TCGA barcodes up to the sample/vial field (16 characters total, e.g. TCGA-AO-A0JJ-01A)

See Also

TCGA.BRCA.Demo, TCGA.BRCA.ExpressionData

TCGA.GBM.Demo

A subset of the glioblastoma multiforme (GBM) data from TCGA.

Description

This dataset is loaded by calling `data(TCGA.GBM.Demo)`.

Details

The loaded data consists of two related parts:

- A matrix of gene expression data.
- A vector containing the TP53 mutation status of each sample in the matrix.

See Also

TCGA.GBM.ExpressionData, TCGA.GBM.TP53MutationData

TCGA.GBM.EXPR	<i>Glioblastoma Multiforme (GBM) microarray expression data from TCGA</i>
---------------	---

Description

Load using `data('TCGA.GBM.EXPR')`.

Format

A numeric data matrix with 2000 rows and 547 columns.

- The data was generated using microarray platforms.
- Row labels are gene symbols (e.g. KRT19). The NG-CHM label type is `bio.gene.hugo`.
- Column labels are TCGA barcodes up to the center field (28 characters total, e.g. TCGA-02-0001-01C-01R-0177-01). The NG-CHM label type is `bio.tcga.barcode.sample.vial.portion.analyte.aliquot`.
- Data has been log-transformed (min 2.196606, max 14.41321).
- The data has no column labels in common with the data in `TCGA.GBM.ExpressionData` (as expected), but at the participant level (first 12 characters) there are 158 columns in common and at the vial level (first 16 characters) there are 152 in common. There are 1098 genes in common. This permits several types of NG-CHMs integrating the two data sets.

TCGA.GBM.ExpressionData	<i>Glioblastoma Multiforme (GBM) RNASeq expression data from TCGA</i>
-------------------------	---

Description

Load using `data('TCGA.GBM.Demo')`.

Format

A numeric data matrix with 3540 rows and 169 columns.

- Row labels are gene symbols (e.g. SYK). The NG-CHM label type is `bio.gene.hugo`.
- Column labels are TCGA barcodes up to the center field (28 characters total, e.g. TCGA-06-0178-01A-01R-1849-01). The NG-CHM label type is `bio.tcga.barcode.sample.vial.portion.analyte.aliquot`.
- Data has been log-transformed and row centered (min -6.373672, max 9.701261).
- This data set and `TCGA.BRCA.ExpressionData` have 1225 genes (rows) in common.
- See `TCGA.GBM.EXPR` for commonalities with that data set.

See Also

`TCGA.GBM.Demo`, `TCGA.GBM.TP53MutationData`, `TCGA.GBM.EXPR`

TCGA.GBM.TP53MutationData

TP53 mutation data for TCGA glioblastoma multiforme (GBM) samples

Description

Load using `data('TCGA.GBM.Demo')`.

Format

A length 169 character vector.

- Each element of the vector is either "WT" or "MUT".
- Element names are TCGA barcodes up to the center field (28 characters total, e.g. TCGA-06-0178-01A-01R-1849-01).

See Also

TCGA.GBM.Demo, TCGA.GBM.ExpressionData