

# R documentation

of all in ‘RTCGA/man’

May 5, 2015

## R topics documented:

RTCGA-package . . . . .	1
availableDataSets . . . . .	2
availableDates . . . . .	3
checkDataSetsAvailability . . . . .	3
checkGenesNamesAvailability . . . . .	4
downloadTCGA . . . . .	5
infoTCGA . . . . .	6
mergeTCGA . . . . .	6
read.clinical . . . . .	7
<b>Index</b>	<b>8</b>

---

RTCGA-package	<i>The Cancer Genome Atlas data integration</i>
---------------	---

---

## Description

The Cancer Genome Atlas (TCGA) Data Portal provides a platform for researchers to search, download, and analyze data sets generated by TCGA. It contains clinical information, genomic characterization data, and high level sequence analysis of the tumor genomes. The key is to understand genomics to improve cancer care. RTCGA package offers download and integration of the variety and volume of TCGA data using patient barcode key, what enables easier data possession. This may have an beneficial influence on impact on development of science and improvement of patients' treatment. Furthermore, RTCGA package transforms TCGA data to form which is convenient to use in R statistical package. Those data transformations can be a part of statistical analysis pipeline which can be more reproducible with RTCGA

## Details

For more detailed information visit **RTCGA** wiki on [Github](#).

## Author(s)

Marcin Kosinski [aut, cre] < m.p.kosinski@gmail.com >  
Przemysław Biecek [aut] < przemyslaw.biecek@gmail.com >

**See Also**

Other RTCGA: [availableDataSets](#); [availableDates](#); [availableGenesNames](#), [checkGenesNamesAvailability](#); [checkDataSetsAvailability](#); [downloadTCGA](#); [infoTCGA](#); [mergeTCGA](#); [read.clinical](#)

---

availableDataSets	<i>TCGA datasets' names</i>
-------------------	-----------------------------

---

**Description**

Enables to check TCGA datasets' names for current release date and cohort.

**Usage**

```
availableDataSets(cancerType, date = NULL)
```

**Arguments**

cancerType	A character of length 1 containing abbreviation (Cohort code) of types of cancers to check for available datasets' names on <a href="http://gdac.broadinstitute.org/">http://gdac.broadinstitute.org/</a> .
date	A NULL or character specifying from which date datasets' names should be checked. By default (date = NULL) the newest available date is used. All available dates can be checked on <a href="http://gdac.broadinstitute.org/runs/">http://gdac.broadinstitute.org/runs/</a> or by using <a href="#">availableDates</a> function. Required format "YYYY-MM-DD".

**Value**

A vector of available datasets' names to pass to the [downloadTCGA](#) function.

**See Also**

Other RTCGA: [RTCGA-package](#); [availableDates](#); [availableGenesNames](#), [checkGenesNamesAvailability](#); [checkDataSetsAvailability](#); [downloadTCGA](#); [infoTCGA](#); [mergeTCGA](#); [read.clinical](#)

**Examples**

```
## Not run:
availableDataSets( "BRCA" )
availableDataSets( "OV", availableDates()[5] ) # error

## End(Not run)
```

---

availableDates	<i>TCGA datasets' releases dates</i>
----------------	--------------------------------------

---

**Description**

Enables to check dates of TCGA datasets' releases.

**Usage**

```
availableDates()
```

**Value**

A vector of available dates to pass to the [downloadTCGA](#) function.

**See Also**

Other RTCGA: [RTCGA-package](#); [availableDataSets](#); [availableGenesNames](#), [checkGenesNamesAvailability](#); [checkDataSetsAvailability](#); [downloadTCGA](#); [infoTCGA](#); [mergeTCGA](#); [read.clinical](#)

**Examples**

```
## Not run:
availableDates()

## End(Not run)
```

---

checkDataSetsAvailability	<i>TCGA datasets' names availability</i>
---------------------------	--

---

**Description**

Enables to check TCGA datasets' names availability for current release date and cancer type.

**Usage**

```
checkDataSetsAvailability(cancerTypes, pattern = "Merge_Clinical.Level_1",
  date = NULL)
```

**Arguments**

cancerTypes	A character vector containing abbreviation (Cohort code) of types of cancers to check for availability of datasets' name on <a href="http://gdac.broadinstitute.org/">http://gdac.broadinstitute.org/</a> .
pattern	A character vector of length 1 containing a part of a dataset's name to be checked for availability for current date parameter. By default phrase "Merge_Clinical.Level_1" is checked.
date	A NULL or character specifying from which date datasets' names should be checked for availability. By default (date = NULL) the newest available date is used. All available dates can be checked on <a href="http://gdac.broadinstitute.org/runs/">http://gdac.broadinstitute.org/runs/</a> or by using <a href="#">availableDates</a> function. Required format "YYYY-MM-DD".

**Value**

A vector of available datasets names to pass to the [downloadTCGA](#) function.

**See Also**

Other RTCGA: [RTCGA-package](#); [availableDataSets](#); [availableDates](#); [availableGenesNames](#), [checkGenesNamesAvailability](#); [downloadTCGA](#); [infoTCGA](#); [mergeTCGA](#); [read.clinical](#)

**Examples**

```
## Not run:
checkDataSetsAvailability( "BRCA" )
checkDataSetsAvailability( c("BRCA", "OV") )
checkDataSetsAvailability( "BRCA", "Mutation_Packager_Calls.Level" )

## End(Not run)
```

---

checkGenesNamesAvailability

*TCGA genes' names and availability in Merge\_rnaseqv2\_\_... dataset.*

---

**Description**

availableGenesNames returns all available genes' names from genes' expressions dataset, where checkGenesNamesAvailability checks whether genes specified in genes are available in Merge\_rnaseqv2\_\_illumina (genes' expressions) dataset.

**Usage**

```
checkGenesNamesAvailability(rnaseqDir, genes)

availableGenesNames(rnaseqDir)
```

**Arguments**

rnaseqDir	A directory to a cancerType.rnaseqv2__illumina_hiseq_rnaseqv2__unc_edu__Level_3__RSEM file
genes	A character - which genes to check for availability in a dataset.

**Value**

A vector containing genes' names that matched existing names.

**See Also**

Other RTCGA: [RTCGA-package](#); [availableDataSets](#); [availableDates](#); [checkDataSetsAvailability](#); [downloadTCGA](#); [infoTCGA](#); [mergeTCGA](#); [read.clinical](#)

Other RTCGA: [RTCGA-package](#); [availableDataSets](#); [availableDates](#); [checkDataSetsAvailability](#); [downloadTCGA](#); [infoTCGA](#); [mergeTCGA](#); [read.clinical](#)

## Examples

```
## Not run:
  checkGenesNamesAvailability( rnaseqDir, "TP53" )

## End(Not run)
```

---

downloadTCGA

*Download TCGA data*


---

## Description

Enables to download TCGA data from specified dates of releases of concrete Cohorts of cancer types. Pass a name of required dataset to the `dataSet` parameter. By default the Merged Clinical `dataSet` is downloaded (value `dataSet = "Merge_Clinical.Level_1"`) from the newest available date of release.

## Usage

```
downloadTCGA(cancerTypes, dataSet = "Merge_Clinical.Level_1", destDir,
  date = NULL)
```

## Arguments

<code>cancerTypes</code>	A character vector containing abbreviations (Cohort code) of types of cancers to download from <a href="http://gdac.broadinstitute.org/">http://gdac.broadinstitute.org/</a> .
<code>dataSet</code>	A part of the name of <code>dataSet</code> to be downloaded from <a href="http://gdac.broadinstitute.org/runs/">http://gdac.broadinstitute.org/runs/</a> . By default the Merged Clinical <code>dataSet</code> is downloaded (value <code>dataSet = "Merge_Clinical.Level_1"</code> ). Available datasets' names can be checked using <a href="#">availableDataSets</a> function.
<code>destDir</code>	A character specifying a directory into which <code>dataSets</code> will be downloaded.
<code>date</code>	A NULL or character specifying from which date <code>dataSets</code> should be downloaded. By default ( <code>date = NULL</code> ) the newest available date is used. All available dates can be checked on <a href="http://gdac.broadinstitute.org/runs/">http://gdac.broadinstitute.org/runs/</a> or by using <a href="#">availableDates</a> function. Required format "YYYY-MM-DD".

## See Also

Other RTCGA: [RTCGA-package](#); [availableDataSets](#); [availableDates](#); [availableGenesNames](#), [checkGenesNamesAvailability](#); [checkDataSetsAvailability](#); [infoTCGA](#); [mergeTCGA](#); [read.clinical](#)

## Examples

```
## Not run:

dir.create( "hre")

downloadTCGA( cancerTypes = "BRCA", dataSet = "miR_gene_expression",
  destDir = "hre/" )

downloadTCGA( cancerTypes = c("BRCA", "OV"), destDir = "hre/" )

## End(Not run)
```

---

infoTCGA

*Information about cohorts from TCGA project*


---

### Description

Function restores codes and counts for each cohort from TCGA project.

### Usage

```
infoTCGA()
```

### Value

A list with a tabular information from <http://gdac.broadinstitute.org/>.

### See Also

Other RTCGA: [RTCGA-package](#); [availableDataSets](#); [availableDates](#); [availableGenesNames](#), [checkGenesNamesAvailability](#); [checkDataSetsAvailability](#); [downloadTCGA](#); [mergeTCGA](#); [read.clinical](#)

---

mergeTCGA

*Merge Clinical data with genes' Mutations and Expressions data*


---

### Description

mergeTCGA enables to

### Usage

```
mergeTCGA(clinicalDir, rnaseqDir = NULL, mutationDir = NULL, genes)
```

### Arguments

**clinicalDir** A directory to a cancerType.clin.merged.txt file. cancerType might be BRCA, OV etc. Can be checked using [infoTCGA](#) function.

**rnaseqDir** A directory to a cancerType.rnaseqv2\_\_illuminahisec\_rnaseqv2\_\_unc\_edu\_\_Level\_3\_\_RSEM file, which is a set with gene's Expressions.

**mutationDir** A directory to a Mutation\_Packager\_Calls.Level1 folder where are genes' Mutations files.

**genes** For rnaseqDir - which genes' expressions to merge with clinical data in clinicalDir. For mutationDir which gene's mutations to merge with clinical data in clinicalDir.

### Value

A cancerType.clin.merged.txt file is updated with newline containing informations about genes passed to genes argument.

**Note**

Original cancerType.clin.merged.txt file will be changed after performing merge operation.  
Only one of rnaseqDir and mutationDir can be used at a time.

**See Also**

Other RTCGA: [RTCGA-package](#); [availableDataSets](#); [availableDates](#); [availableGenesNames](#),  
[checkGenesNamesAvailability](#); [checkDataSetsAvailability](#); [downloadTCGA](#); [infoTCGA](#); [read.clinical](#)

---

read.clinical	<i>Read from txt fo,e</i>
---------------	---------------------------

---

**Description**

TODO

**Usage**

```
read.clinical(clinicalDir, ...)
```

**Arguments**

clinicalDir	A directory to a cancerType.clin.merged.txt file. cancerType might be BRCA, OV etc.
-------------	---

**Value**

A data.frame with clinical data.

**See Also**

Other RTCGA: [RTCGA-package](#); [availableDataSets](#); [availableDates](#); [availableGenesNames](#),  
[checkGenesNamesAvailability](#); [checkDataSetsAvailability](#); [downloadTCGA](#); [infoTCGA](#); [mergeTCGA](#)

# Index

availableDataSets, [2](#), [2](#), [3–7](#)  
availableDates, [2](#), [3](#), [3](#), [4–7](#)  
availableGenesNames, [2–7](#)  
availableGenesNames  
    (checkGenesNamesAvailability),  
    [4](#)  
  
checkDataSetsAvailability, [2](#), [3](#), [3](#), [4–7](#)  
checkGenesNamesAvailability, [2–4](#), [4](#), [5–7](#)  
  
downloadTCGA, [2–4](#), [5](#), [6](#), [7](#)  
  
infoTCGA, [2–6](#), [6](#), [7](#)  
  
mergeTCGA, [2–6](#), [6](#), [7](#)  
  
read.clinical, [2–7](#), [7](#)  
RTCGA-package, [1](#)