

# Package ‘SomaticCancerAlterations’

April 12, 2018

**Type** Package

**Title** Somatic Cancer Alterations

**Version** 1.14.0

**Author** Julian Gehring (EMBL Heidelberg)

**Maintainer** Julian Gehring <jg-bioc@gmx.com>

**Imports** GenomicRanges, exomeCopy, stringr, IRanges, S4Vectors

**Depends** R (>= 3.0.0)

**Suggests** testthat, ggbio, ggplot2, knitr

**VignetteBuilder** knitr

**Description** Collection of somatic cancer alteration datasets

**License** GPL-3

**LazyLoad** yes

**biocViews** ExperimentData, Genome, CancerData, Project1000genomes, NCI

**Encoding** UTF-8

**NeedsCompilation** no

## R topics documented:

SomaticCancerAlterations-package . . . . .	1
SomaticCancerAlterations-data . . . . .	2
SomaticCancerAlterations-functions . . . . .	3
SomaticCancerAlterations-helper-functions . . . . .	5
<b>Index</b>	<b>6</b>

---

SomaticCancerAlterations-package  
*Somatic Cancer Alterations*

---

## Description

A collection of studies with somatic cancer alterations.

## Details

The 'SomaticCancerAlterations' package provides a collection of the mutational calls for different cancer studies, aiming for a tight integration with R and Bioconductor. At the moment, this covers somatic single nucleotide variants (SNVs) and indels for several TCGA studies. Over time, this collection will be updated to include additional studies.

The Cancer Genome Atlas (TCGA) is a consortium effort to analyze a variety of tumor types at multiple levels, including gene expression, methylation, copy number alterations and somatic mutations [<http://cancergenome.nih.gov>]. Most of this data is publically available, and offers a rich resource in understanding cancer-related data sets and experiments [<https://wiki.nci.nih.gov/display/TCGA/TCGA+Home>].

## Author(s)

Julian Gehring (EMBL Heidelberg), with suggestions and contributions from Bernd Fischer  
Maintainer: Julian Gehring <[julian.gehring@embl.de](mailto:julian.gehring@embl.de)>

## References

<https://tcga-data.nci.nih.gov/tcga/>  
<https://tcga-data.nci.nih.gov/docs/publications/>

## See Also

[SomaticCancerAlterations-functions](#)  
[SomaticCancerAlterations-data](#)

---

SomaticCancerAlterations-data  
*Somatic Cancer Alterations Data Sets*

---

## Description

Data sets with studies of somatic cancer alterations.

## Details

Currently, the data sets include somatic mutations (SNVs and indels) of the publically available TCGA studies. Further studies will be included in the future.

For details on the TCGA mutation calls, please see the TCGA documentation <https://wiki.nci.nih.gov/display/TCGA/TCGA+Data+Primer>. Please note that the mutational calling was performed by different centers. The original '\*.maf' files can be obtained from the TCGA data portal [https://tcga-data.nci.nih.gov/tcgafiles/ftp\\_auth/distro\\_ftpusers/anonymous/tumor/](https://tcga-data.nci.nih.gov/tcgafiles/ftp_auth/distro_ftpusers/anonymous/tumor/).

After importing, the data is automatically processed to improve consistency across studies and remove common artifacts in the annotation. The most obvious changes are:

- Only studies cleared for unrestricted usage are included (see <http://cancergenome.nih.gov/publications/publicationguidelines>).
- Only studies processed by the Broad Institute are considered at the moment.

- All variants located on the mitochondrion are mapped to the sequence identifier 'MT' (note that the original files contain both 'MT' and 'M' mixed).
- All locations are now consistent with the 1000genomes reference sequence (NCBI37 coordinates).

About the TCGA data:

“All data generated by The Cancer Genome Atlas (TCGA) Research Network are made open to the public through the Data Coordinating Center and the TCGA Data Portal.” For details on the usage of the data, please have a look at <http://cancergenome.nih.gov/abouttcga/policies/publicationguidelines>.

## Value

A 'GRanges' object for each study. Each row corresponds to a somatic variant reported in the respective study, with the coordinates referring to the location on the reference genome.

## References

<https://tcga-data.nci.nih.gov/tcga/>

## See Also

[SomaticCancerAlterations-functions](#)

[SomaticCancerAlterations-package](#)

## Examples

```
all_datasets = scaListDatasets()
gr1 = scaLoadDatasets(all_datasets[1])
```

---

SomaticCancerAlterations-functions

*SomaticCancerAlterations functions*

---

## Description

Functions to interact with the 'SomaticCancerAlterations' datasets.

## Usage

```
scaMetadata()
```

```
scaListDatasets()
```

```
scaLoadDatasets(names, merge = FALSE)
```

```
mutationDensity(x, binSize = 1e6, chrs = seqnames(x))
```

**Arguments**

names	Character vector with data set name(s) to load. The names are the same as returned from 'scaListDatasets()'.
merge	Logical. Should the results be a merged 'GRanges' object with data of all studies (TRUE), or a 'GenomicRangesList' with one list element per dataset (FALSE [default]).
x	GRanges object
binSize	Integer with the bin size used for computing the density of variants.
chrs	Character vector with the chromosomes/seqlevels to include (by default, all chromosomes are used.)

**Details**

The 'sca\_load\_dataset' function makes it easy to load the data of a study directly into a variable.

The 'mutationDensity' function computes the density of events in a binned-manner.

**Value**

scaMetadata	A data frame summarizing the available data sets, with rows corresponding to data sets.
scaListDatasets	A character vector with all available study names.
sca_load_dataset	A GenomicRangesList or GRanges objects, depending on the 'merge' argument.
mutationDensity	A GRanges object, with ranges corresponding to bins along the genome. The columns 'counts' and 'density' refer to the number and the fraction (given per 1Mbp) of alterations, respectively.

**References**

<https://tcga-data.nci.nih.gov/tcga/>

<https://wiki.nci.nih.gov/display/TCGA/TCGA+Data+Primer>

[https://tcga-data.nci.nih.gov/tcgafiles/ftp\\_auth/distro\\_ftpusers/anonymous/tumor/](https://tcga-data.nci.nih.gov/tcgafiles/ftp_auth/distro_ftpusers/anonymous/tumor/)

**See Also**

[SomaticCancerAlterations-data](#)

**Examples**

```
meta_data = scaMetadata()
head(meta_data)

all_datasets = scaListDatasets()
head(all_datasets)

grl = scaLoadDatasets(all_datasets[1], merge = FALSE)

gr = scaLoadDatasets(all_datasets[1], merge = TRUE)
```



# Index

## \*Topic **datasets**

SomaticCancerAlterations-data, [2](#)  
SomaticCancerAlterations-package,

[1](#)

## \*Topic **package**

SomaticCancerAlterations-package,

[1](#)

blca\_tcga

(SomaticCancerAlterations-data),

[2](#)

cesc\_tcga

(SomaticCancerAlterations-data),

[2](#)

gbm\_tcga

(SomaticCancerAlterations-data),

[2](#)

hg2ncbi

(SomaticCancerAlterations-helper-functions),

[5](#)

hnsc\_tcga

(SomaticCancerAlterations-data),

[2](#)

kirc\_tcga

(SomaticCancerAlterations-data),

[2](#)

lgg\_tcga

(SomaticCancerAlterations-data),

[2](#)

luad\_tcga

(SomaticCancerAlterations-data),

[2](#)

lusc\_tcga

(SomaticCancerAlterations-data),

[2](#)

meta (SomaticCancerAlterations-data), [2](#)

mutationDensity

(SomaticCancerAlterations-functions),

[3](#)

ncbi2hg

(SomaticCancerAlterations-helper-functions),

[5](#)

ov\_tcga

(SomaticCancerAlterations-data),

[2](#)

paad\_tcga

(SomaticCancerAlterations-data),

[2](#)

prad\_tcga

(SomaticCancerAlterations-data),

[2](#)

scaListDatasets

(SomaticCancerAlterations-functions),

[3](#)

scaLoadDatasets

(SomaticCancerAlterations-functions),

[3](#)

scaMetadata

(SomaticCancerAlterations-functions),

[3](#)

seqchr

(SomaticCancerAlterations-helper-functions),

[5](#)

skcm\_tcga

(SomaticCancerAlterations-data),

[2](#)

SomaticCancerAlterations

(SomaticCancerAlterations-package),

[1](#)

SomaticCancerAlterations-data, [2](#)

SomaticCancerAlterations-functions, [3](#)

SomaticCancerAlterations-helper-functions,

[5](#)

SomaticCancerAlterations-package, [1](#)

stad\_tcga

(SomaticCancerAlterations-data),

[2](#)

thca\_tcga

(SomaticCancerAlterations-data),

[2](#)

ucec\_tcga  
    (SomaticCancerAlterations-data),  
    [2](#)