Package 'SomaticCancerAlterations'

April 9, 2026

Type Package

Title Somatic Cancer Alterations

Version 1.46.0

Author Julian Gehring

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

Imports GenomicRanges, 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

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

git_branch RELEASE_3_22

git_last_commit 22e826f

git_last_commit_date 2025-10-29

Repository Bioconductor 3.22

Date/Publication 2026-04-09

Contents

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 colletion 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, with suggestions and contributions from Bernd Fischer

Maintainer: Julian Gehring

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/.

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 Instutite are considered at the moment.
- All variants located on the mitochondrium 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 refering 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()
grl = scaLoadDatasets(all_datasets[1])
```

SomaticCancerAlterations-functions
SomaticCancerAlterations functions

Description

Functions to interact with the 'SomaticCancerAlterations' datasets.

Usage

```
scaMetadata()

scaListDatasets()

scaLoadDatasets(names, merge = FALSE)
```

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 'GRangesList' with one list element per dataset (FALSE [default]).

Details

The 'sca_load_dataset' function makes it easy to load the data of a study directly into a variable.

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 GRangesList or GRanges objecs, depending on the 'merge' argument.

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/

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)
```

SomaticCancerAlterations-helper-functions
SomaticCancerAlterations helper functions

Description

Functions to help with common analysis tasks.

Usage

```
ncbi2hg(x)

hg2ncbi(x)

seqchr(x)
```

Arguments

x	A GRanges object

Details

These functions provide useful wrappers for converting between genomic coordinate systems and classes.

Value

ncbi2hg, hg2ncbi	A GRanges object.
seqchr	A character vector with the 'seqnames'.

Index