Package ‘PCAWG7’

August 8, 2021

Title Repository of data from COSMIC and 'The repertoire of Mutational Signatures in Human Cancer'

Version 0.1.1

Description Contains data from The COSMIC
web site https://cancer.sanger.ac.uk/cosmic/signatures/index.tt
and from the paper by Alexandrov, Kim, Haradhvala, Huang et al.,
'The repertoire of Mutational Signatures in Human Cancer'. Please see ?PCAWG7.
https://doi.org/10.1038/s41586-020-1943-3. The funny name
comes from the fact that this paper was generated by
Working Group 7 of the Pan Cancer Analysis of Whole Genomes
(PCAWG) consortium. Soem of the data were then placed on the COSMIC
web site and subsequently updated.

License GPL-3

Language en-US

Encoding UTF-8

LazyData true

Depends R (>= 3.5),

RoxygenNote 7.1.1

URL `https://github.com/steverozen/PCAWG7`

BugReports `https://github.com/steverozen/PCAWG7/issues`

Imports ICAMS (>= 2.3.5.9002)

Remotes github::steverozen/ICAMS@master

Suggests usethis,
testthat (>= 3.0.0)

Config/testthat/edition 3

R topics documented:

`CancerTypes` *Return a character vector of some common cancer types*

Description

Return a character vector of some common cancer types

Usage

```
CancerTypes()
```

Examples

```
CancerTypes()[1:5]
```

`COSMIC.v3.0` *PCAWG7 SigProfiler reference signatures.*

Description

PCAWG7 SigProfiler reference signatures.

Usage

```
COSMIC.v3.0
```

Format

A list with one element, `signature`, which in turn is a list with elements:

genome A list with the elements:

- **SBS96** Strand-agnostic single-base substitutions in trinucleotide context.
- **SBS192** Transcriptionally stranded single-base substitutions in trinucleotide context.
- **DBS78** Strand-agnostic doublet-base substitutions.
- **ID** Strand-agnostic indels.

exome A list with the elements:

- **SBS96** As above, for exome count signatures, which look different than genome count signatures, because of differences in trinucleotide frequencies in exomes versus whole genomes. These were signatures that were extracted from exome data in the PCAWG7 paper, not simple adjustment of the `genome` signatures for exome trinucleotide abundances.

Source

Subdirectories of https://www.synapse.org/#!Synapse:syn12009743, 2019 Oct 09, populated by `data-raw/populate.variable.siganture.R`.

Examples

```
SBS96.sigs <- COSMIC.v3.0$signature$genome$SBS96
```

`exposure` *PCAWG7 SigProfiler signature assignments (numbers of mutations due to each signature in each tumor).*

Description

PCAWG7 SigProfiler signature assignments (numbers of mutations due to each signature in each tumor).

Usage

```
exposure
```

Format

A list with the elements:

- **PCAWG** A list with the elements:
  - **SBS96** Strand-agnostic single-base substitutions in trinucleotide context.
  - **DBS78** Strand-agnostic doublet-base substitutions.
  - **ID** Strand-agnostic indels. These are signature assignments for the PCAWG platinum genomes.
- **TCGA** A list with the elements:
  - **SBS96** As above.
  - **ID** As above. These are signature assignments for the TCGA exomes.
- **other.genome** A list with the element:
  - **SBS96** As above. This contains signature assignments for non-TCGA genomes.
- **other.exome** A list with the element:
  - **SBS96** As above. This contains signature assignments for non-TCGA exomes.

Source

Files of https://www.synapse.org/#!Synapse:syn12009743, 2019 Oct 09, populated by `data-raw/sig.profiler.sures/load.package.variable.exposure.R`.

Examples

```
SBS96.exposure <- exposure$PCAWG$SBS96
```

exposure.stats *Exposure statistics from the PCAWG7 paper*

Description

Exposure statistics from the PCAWG7 paper

Usage

```
exposure.stats
```

Format

A list with one element, `PCAWG`, which has the sub-elements `SBS96`, `DBS78`, `ID` with statistics for the corresponding mutation types by cancer type. I.e. each element has a sub-element for each cancer type, and this element is a data.frame with one row for each signature and columns `mean.of.those.present` (the mean number of mutations for those tumors that have the mutation) and `proportion.present` (the proportion of tumors in which the signature is present).

Source

Computed from other package variables using `GatherPCAWG7ExposureStatsSBS96`.

Examples

```
exposure.stats$PCAWG$SBS96$`Biliary-AdenoCA`[1:3, ]
```

GetEtiology *Get the proposed etiology of a signature*

Description

Get the proposed etiology of a signature

Usage

```
GetEtiology(mutation.type, sig.id)
```

Arguments

`mutation.type`
character string, one of SBS96, SBS192, DBS78, ID

`sig.id` character vector with signature signature ids, e.g. `c("SBS3","foo")`.

Value

A character vector of the same length as `sig.id`, each element of which is the etiology of the corresponding signature, if known, or else the empty string.

Examples

```
GetEtiology(mutation.type = "ID", sig.id = c("ID1", "foo", "ID3"))
```

map_SP_ID_to_tumor_type

Given PCAWG "SP" IDs (eg. SP123958) return either the "full" IDs (Kidney-ChRCC::SP123958) or just the tumor type (Kidney-ChRCC)

Description

Given PCAWG "SP" IDs (eg. SP123958) return either the "full" IDs (Kidney-ChRCC::SP123958) or just the tumor type (Kidney-ChRCC)

Usage

```
map_SP_ID_to_tumor_type(SP.IDs, merge = TRUE)
```

Arguments

`SP.IDs`	A character vector of PCAWG "SP" IDs.
`merge`	If `TRUE` return a parallel vector of <tumor_type>::<SP_ID>; otherwise just <tumor_type>.

Details

Fails with an "subscript out of bounds" error if any of the elements of `SP.IDs` is unknown.

Examples

```
map_SP_ID_to_tumor_type(c("SP123958", "SP43633"))
map_SP_ID_to_tumor_type(c("SP123958", "SP43633"), merge = FALSE)
```

PCAWG.sample.id

Vectors of the PCAWG `tumor_wgs_icgc_specimin_id`*s.*

Description

Note that the PCAWG7 spectra catalogs have 2 sample ids that were blacklisted after the mutational signature analysis was underway. The blacklisted samples are `SP116419` and `SP116883`, which are in `PCAWG.sample.id$black`.

Usage

```
PCAWG.sample.id
```

Format

A list with the elements:

white Whitelisted IDs

grey Greylisted IDs

black Blacklisted IDs

Source

https://dcc.icgc.org/api/v1/download?fn=/PCAWG/data_releases/latest/release_may2016.v1.4.with_consensus_calls.tsv, 2019 Oct 09

PCAWG7 *PCAWG7: A package of data from 'Repertoire of Mutational Signatures in Human Cancer'*

Description

This is a data package with 3 main package variables: `exposure`, `signature`, and `spectra`.

Details

There are also PDF plots of the signatures in `data-raw/plots/`.

There are also several functions for handling PCAWG identifiers:

* `map_SP_ID_to_tumor_type`

* `SampleIDToCancerType`

* `SplitPCAWGMatrixByTumorType`

* `SplitMatrixBySampleType`

The reference for the data is

Alexandrov, L.B., Kim, J., Haradhvala, N.J. et al. The repertoire of mutational signatures in human cancer. Nature 578, 94-101 (2020). https://doi.org/10.1038/s41586-020-1943-3.

SampleIDToCancerType *Split out the cancer type from the sample ID for PCAWG IDs*

Description

Split out the cancer type from the sample ID for PCAWG IDs

Usage

```
SampleIDToCancerType(PCAWGID)
```

Arguments

`PCAWGID`	A character vector of PCAWG IDs of the form <cancer.type>::<sample.id>.

Value

A character vector parallel to `PCAWGID` containing only the <cancer.type> strings.

Examples

```
cancer.type <- SampleIDToCancerType("Biliary-AdenoCA::SP117655")
```

SBS96_ID_to_SBS192_ID

Translate SBS96 signature IDs to SBS192 signature IDs by adding "-E" if necessary.

Description

Translate SBS96 signature IDs to SBS192 signature IDs by adding "-E" if necessary.

Usage

```
SBS96_ID_to_SBS192_ID(sig.ids)
```

Arguments

`sig.ids`	Character vector of SBS96 signature IDs.

Value

Character vector of corresponding SBS192 signature IDs; some have "-E" (for exome) post-pended.

Examples

```
SBS96.ids <- c("SBS1", "SBS23", "SBS25")
SBS192.ids <- SBS96_ID_to_SBS192_ID(SBS96.ids)
```

---

signature

Mutational signatures data from COSMIC, the Catalogue Of Somatic Mutations In Cancer, (v3.1 - June 2020)

Description

Mutational signatures data from COSMIC, the Catalogue Of Somatic Mutations In Cancer, (v3.1 - June 2020)

Usage

```
signature
```

Format

A list with a single element, `genome`, which is a list containing:

SBS96 Strand-agnostic single-base substitutions in trinucleotide context.

SBS192 Transcriptionally stranded single-base substitutions in trinucleotide context.

DBS78 Strand-agnostic doublet-base substitutions.

ID Strand-agnostic indels.

Remark

The signatures are all from Human GRCh37 reference genome.

Source

Files downloaded from `https://cancer.sanger.ac.uk/cosmic/signatures/index.tt`, 2021 Feb and saved in `data-raw/COSMIC.v3.1/data/`.
Populated by `data-raw/COSMIC.v3.1/code/generate-COSMIC.v3.1-genome-sigs.R`.

Examples

```
SBS96.sigs <- signature$genome$SBS96
```

`sigs.etiologies` *List of proposed etiologies from PCAWG7 paper, some manually abbreviated and a few summarized from the COSMIC web site.*

Description

List of proposed etiologies from PCAWG7 paper, some manually abbreviated and a few summarized from the COSMIC web site.

Usage

```
sigs.etiologies
```

Format

A list with the elements:

SBS96

SBS192

DBS78

ID

Each list element is a single column matrix with rownames being the signature IDs and values being a short character string description of the proposed etiology.

In general use `GetEtiology`, which handles new signatures without elements in `sigs.etiologies`.

`spectra` *PCAWG7 mutational spectra (catalogs).*

Description

PCAWG7 mutational spectra (catalogs).

Usage

```
spectra
```

Format

A list with the elements:

SBS96 Deprecated.

DBS78 Deprecated.

PCAWG A list with the elements:

- **SBS96** Strand-agnostic single-base substitutions in trinucleotide context.
- **SBS192** Single-base substitutions in transcripts based on the sense strand.
- **SBS1536** Strand-agnostic single-base substitutions in pentanucleotide context.
- **DBS78** Strand-agnostic doublet-base substitutions.
- **ID** Strand-agnostic indels.

TCGA A list with the same elements as the `PCAWG` element.

other.genome A list with the same elements as the `PCAWG` element but with `ID` omitted.

other.exome A list with the same elements as the `PCAWG` element but with `ID` omitted.

Source

Files below `https://www.synapse.org/#!Synapse:syn11801889`, 2019 Oct 09. Populated by `data-raw/spectra/load.package.variable.specra.R`.

Examples

```
SBS96.spectra <- spectra$PCAWG$SBS96
```

`SplitMatrixBySampleType` *Split an exposure matrix or spectrum matrix into a list of matrices, each for a single sample type.*

Description

Split an exposure matrix or spectrum matrix into a list of matrices, each for a single sample type.

Usage

```
SplitMatrixBySampleType(M, sample.type)
```

Arguments

`M` A numerical matrix or data frame or ICAMS catalog in which columns are samples (e.g. tumors) and rows are either mutational signatures (for exposures) or mutation types (for spectra), and, each element is the number of mutations due to a given mutational signature or mutation type in a single sample

`sample.type` A character or numeric vector, each element of which indicates a particular sample type.

Value

Invisibly, the list of exposure or spectrum matrices created by splitting `M` by `sample.type`.

Examples

```
ff <- matrix(1, nrow=3, ncol = 2)
colnames(ff) <- c("sample1", "sample2")
xx <- SplitMatrixBySampleType(ff, c("sample.type.x", "sample.type.y"))
xx
```

SplitPCAWGMatrixByTumorType

Extract tumor type from column names and return the input matrix split by tumor type based on the PCAWG <tumor_type>::<sample_id> convention.

Description

Extract tumor type from column names and return the input matrix split by tumor type based on the PCAWG <tumor_type>::<sample_id> convention.

Usage

```
SplitPCAWGMatrixByTumorType(M)
```

Arguments

`M` A numerical matrix or data frame or ICAMS catalog in which columns are samples (e.g. tumors) and rows are either mutational signatures (for exposures) or mutation types (for spectra), and each element is the number of mutations due to a given mutational signature or mutation type in a single sample. The column names must be of the the form <cancer.type>::<sample.ID>.

Value

Invisibly, the list of exposure matrices or ICAMS catalogs created by splitting `matrix` by the tumor type encoded in the column names.

Examples

```
mm <- SplitPCAWGMatrixByTumorType(spectra$PCAWG$DBS78)
mm[[3]][1:4, 1:5]
```

`TCGA_ID_to_ICGC_ID` *Translate TCGA(The Cancer Genome Atlas) IDs to ICGC(International Cancer Genome Consortium) IDs*

Description

Translate TCGA(The Cancer Genome Atlas) IDs to ICGC(International Cancer Genome Consortium) IDs

Usage

```
TCGA_ID_to_ICGC_ID(tcga.ids)
```

Arguments

`tcga.ids`	Character vector of TCGA IDs.

Value

Character vector of corresponding ICGC IDs. If a corresponding ICGC ID cannot be found, then return an empty string.

Examples

```
tcga.ids <- c("TCGA-AA-A01V", "foo", "TCGA-CA-6717", "bar")
icgc.ids <- TCGA_ID_to_ICGC_ID(tcga.ids)
icgc.ids <- icgc.ids[nzchar(icgc.ids)]
```

Index