Package 'PCAWG7'

February 27, 2021

Title Repository of data from 'Repertoire of Mutational Signatures in Human Cancer'

Version 0.0.3.9006

Description Contains data from Alexandrov, Kim, Haradhvala, Huang et al., 'Repertoire of Mutational Signatures in Human Cancer'. Please see ?PCAWG7. 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. 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.

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

Suggests ICAMS, usethis

R topics documented:

`COSMIC.v3.1` *COSMIC v3.1 data*

Description

COSMIC v3.1 data

Usage

```
COSMIC.v3.1
```

Format

A list with the elements:

signature A list with the 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.

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`.

`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.s

---

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, ]
```

PCAWG.sample.id *Vectors of the PCAWG* tumor_wgs_icgc_specimin_ids.

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

PCAWG.WGS.DBS *Doublet Base Substitution (SBS) spectra (deprecated). Use* spectra$PCAWG$DBS78 *instead.*

Description

Doublet Base Substitution (SBS) spectra (deprecated). Use spectra$PCAWG$DBS78 instead.

Usage

```
PCAWG.WGS.DBS
```

Format

An object of class matrix (inherits from array) with 78 rows and 2780 columns.

PCAWG.WGS.SBS.96 *Single Base Substitution (SBS) spectra in trinucleotide context (deprecated). Use* spectra$PCAWG$SBS96 *instead.*

Description

Single Base Substitution (SBS) spectra in trinucleotide context (deprecated). Use spectra$PCAWG$SBS96 instead.

Usage

```
PCAWG.WGS.SBS.96
```

Format

An object of class matrix (inherits from array) with 96 rows and 2780 columns.

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/.

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.

signature *PCAWG7 SigProfiler reference signatures.*

Description

PCAWG7 SigProfiler reference signatures.

Usage

```
signature
```

Format

A list with the 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.

Source

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

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`.

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

Description

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

Usage

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

Arguments

`M`	A numerical matrix or data frame 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 matrices created by splitting `M` by `sample.type`.

SplitPCAWGMatrixByTumorType *Extract tumor type from column names and return the input matrix split by tumor type.*

Description

Extract tumor type from column names and return the input matrix split by tumor type.

Usage

```
SplitPCAWGMatrixByTumorType(M)
```

Arguments

M	A numerical matrix or data frame 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.

Value

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

Index