Package: protean (via r-universe)

August 27, 2024

Type Package

Title Sequence Profiles of OncoKB Genes

Version 0.1.2

Description A data package of sequence profiles of OncoKB genes. These profiles are obtained via Ensembl's REST API and derived from the pairwise alignment of the human sequence with its orthologs.

License CC BY 4.0

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Imports dplyr, glue, httr2, magrittr, memoise, purrr, readr, rlang, tibble, tidyjson

Suggests here

Depends R (>= 2.10)

Roxygen list(markdown = TRUE)

URL `https://github.com/maialab/protean`

BugReports `https://github.com/maialab/protean/issues`

Repository https://patterninstitute.r-universe.dev

RemoteUrl https://github.com/maialab/protean

RemoteRef HEAD

RemoteSha 0e6628d51e05bbb056404c7c87f65875624c7d76

Contents

`download_gene_list` *Download OncoKB Cancer Gene List*

Description

Download OncoKB Cancer Gene List

Usage

```
download_gene_list(
  path = stop("`path` must be specified"),
  url = oncokb_dwl_url()
)
```

Arguments

`path`	A character string with the file path where the downloaded file is to be saved. Tilde-expansion is performed.
`url`	The URL of the resource providing the OncoKB cancer gene list.

`exported_genes` *Exported genes*

Description

A character vector of genes for which the retrieval of sequence profiles was successful and are hence provided with this package.

Usage

```
exported_genes
```

Format

A character vector.

`fetch_oncokb_genes` *Fetch current OncoKB genes*

Description

`fetch_oncokb_genes()` retrieves the current set of OncoKB genes from an OncoKB's cancer gene list file.

Usage

```
fetch_oncokb_genes(file = oncokb_dwl_url())
```

Arguments

`file`	A URL or a file path to the source providing the cancer gene list file. By default it will automatically download cancerGeneList.tsv from OncoKB website.

Value

A character vector of gene names.

Examples

```
fetch_oncokb_genes()
```

`get_profile` *Get sequence profiles*

Description

This function retrieves pairwise alignments between the human sequence queried in `symbol` and each of its orthologs via Ensembl's REST API `homology/symbol/:species/:symbol` endpoint. Then, from these alignments, sequence profiles are derived.

Usage

```
get_profile(symbol, simplify = TRUE)
```

Arguments

`symbol`	A character vector of HUGO gene symbols.
`simplify`	Should the result be simplified if only one gene symbol is queried. If `TRUE`, then in the case only one gene symbol is queried the result is not a list of one tibble, but the tibble itself.

Value

A list of tibbles, one for each gene symbol queried, with the following columns:

`timestamp`	Date and time of the download from Ensembl.
`human_prot_id`	Ensembl identifier of the human protein sequence.
`ortho_prot_id`	Ensembl identifier of the ortholog protein sequence.
`ortho_species`	Species name of the ortholog sequence.
`human_align_seq`	In the context of pairwise alignment between the human sequence and one of its orthologs, this is the aligned human sequence.
`ortho_align_seq`	In the context of pairwise alignment between the human sequence and one of its orthologs, this is the aligned ortholog sequence.
`human_ortho_perc_id`	Percentage of the human sequence matching the sequence of the ortholog.
`ortho_human_perc_id`	Percentage of the orthologous sequence matching the human sequence.
`human_profile_id`	Human protein sequence.
`ortho_profile_seq`	Orthologous sequence stripped off of the alignment positions which correspond to gaps in the human sequence.

`missing_genes` *Missing genes*

Description

A character vector of genes for which the retrieval of sequence profiles was not successful and are therefore absent.

Usage

```
missing_genes
```

Format

A character vector.

`oncokb_genes` *OncoKB genes*

Description

A character vector of OncoKB genes used as query to retrieve the sequence profiles bundled with this package.

Usage

```
oncokb_genes
```

Format

A character vector.

`profile_path` *Get the path to a sequence profile*

Description

protean comes bundled with a number of sequence profile files in its `inst/profiles` directory. This function make them easy to access by returning the local path to them.

Usage

```
profile_path(file = NULL)
```

Arguments

`file`	Name of file or gene symbol. If `NULL`, the profile files will be listed.

Examples

```
# Retrieve the path to the sequence profile of the TP53 protein
# Using the gene symbol
profile_path("TP53")

# Using the file name
profile_path("TP53.csv.gz")

# List all profile files
profile_path()
```

`read_profile` *Read a sequence profile*

Description

Read a sequence profile

Usage

```
read_profile(file = stop("`file` must be specified"), sort = TRUE)
```

Arguments

`file`	A path to a sequence profile file.
`sort`	Whether to sort the sequences by the variable `human_ortho_perc_id`, from highest (most similar to human) to lowest (most distant from human).

Value

A tibble of 10 variables:

`timestamp`	Date and time of the download from Ensembl.
`human_prot_id`	Ensembl identifier of the human protein sequence.
`ortho_prot_id`	Ensembl identifier of the ortholog protein sequence.
`ortho_species`	Species name of the ortholog sequence.
`human_align_seq`	In the context of pairwise alignment between the human sequence and one of its orthologs, this is the aligned human sequence.
`ortho_align_seq`	In the context of pairwise alignment between the human sequence and one of its orthologs, this is the aligned ortholog sequence.
`human_ortho_perc_id`	Percentage of the human sequence matching the sequence of the ortholog.
`ortho_human_perc_id`	Percentage of the orthologous sequence matching the human sequence.
`human_profile_id`	Human protein sequence.
`ortho_profile_seq`	Orthologous sequence stripped off of the alignment positions which correspond to gaps in the human sequence.

Examples

```
read_profile(profile_path("TP53"))
```

Index