

Package: driveR (via r-universe)

August 21, 2024

Title Prioritizing Cancer Driver Genes Using Genomics Data

Version 0.4.1.9001

Maintainer Ege Ulgen <egeulgen@gmail.com>

Description Cancer genomes contain large numbers of somatic alterations but few genes drive tumor development. Identifying cancer driver genes is critical for precision oncology. Most of current approaches either identify driver genes based on mutational recurrence or using estimated scores predicting the functional consequences of mutations. 'driveR' is a tool for personalized or batch analysis of genomic data for driver gene prioritization by combining genomic information and prior biological knowledge. As features, 'driveR' uses coding impact metaprediction scores, non-coding impact scores, somatic copy number alteration scores, hotspot gene/double-hit gene condition, 'phenolyzer' gene scores and memberships to cancer-related KEGG pathways. It uses these features to estimate cancer-type-specific probability for each gene of being a cancer driver using the related task of a multi-task learning classification model. The method is described in detail in Ulgen E, Sezerman OU. 2021. driveR: driverR: a novel method for prioritizing cancer driver genes using somatic genomics data. BMC Bioinformatics <doi:10.1186/s12859-021-04203-7>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

URL <https://egeulgen.github.io/driveR/>,
<https://github.com/egeulgen/driveR/>

BugReports <https://github.com/egeulgen/driveR/issues>

Imports caret, randomForest, GenomicRanges, GenomeInfoDb, GenomicFeatures, TxDb.Hsapiens.UCSC.hg19.knownGene, TxDb.Hsapiens.UCSC.hg38.knownGene, S4Vectors, org.Hs.eg.db, rlang,

Depends R (>= 4.0)

Suggests testthat, covr, knitr, rmarkdown

VignetteBuilder knitr

Repository <https://egeulgen.r-universe.dev>

RemoteUrl <https://github.com/egeulgen/driver>

RemoteRef HEAD

RemoteSha 978fb132d8c27228a6068a4604fa86c669369f76

Contents

create_features_df	2
create_gene_level_scna_df	5
create_noncoding_impact_score_df	6
create_SCNA_score_df	7
determine_double_hit_genes	8
determine_hotspot_genes	8
driveR	9
example_cohort_features_table	10
example_cohort_scna_table	11
example_features_table	12
example_gene_scna_table	13
example_scna_table	14
KEGG_cancer_pathways_descriptions	14
MTL_submodel_descriptions	15
predict_coding_impact	15
prioritize_driver_genes	16
specific_thresholds	18
Index	19

create_features_df *Create Data Frame of Features for Driver Gene Prioritization*

Description

Create Data Frame of Features for Driver Gene Prioritization

Usage

```

create_features_df(
  annovar_csv_path,
  scna_segs_df,
  scna_genes_df,
  phenolyzer_annotated_gene_list_path,
  batch_analysis = FALSE,
  prep_phenolyzer_input = FALSE,
  build = "GRCh37",
  log2_ratio_threshold = 0.25,
  gene_overlap_threshold = 25,
  MCR_overlap_threshold = 25,
  hotspot_threshold = 5L,
  log2_hom_loss_threshold = -1,
  verbose = TRUE,
  na.string = "."
)

```

Arguments

annovar_csv_path path to 'ANNOVAR' csv output file

scna_segs_df the SCNA segments data frame. Must contain:
chr chromosome the segment is located in
start start position of the segment
end end position of the segment
log2ratio \log_2 ratio of the segment

scna_genes_df data frame of gene-level SCNAs (can be output of [create_gene_level_scna_df](#))

phenolyzer_annotated_gene_list_path path to 'phenolyzer' 'annotated_gene_list' file

batch_analysis boolean to indicate whether to perform batch analysis (TRUE, default) or personalized analysis (FALSE). If TRUE, a column named 'tumor_id' should be present in both the ANNOVAR csv and the SCNA table.

prep_phenolyzer_input boolean to indicate whether or not to create a vector of genes for use as the input of 'phenolyzer' (default = FALSE). If TRUE, the features data frame is not created and instead the vector of gene symbols (union of all genes for which scores are available) is returned.

build genome build for the SCNA segments data frame (default = 'GRCh37')

log2_ratio_threshold the \log_2 ratio threshold for keeping high-confidence SCNA events (default = 0.25)

gene_overlap_threshold the percentage threshold for the overlap between a segment and a transcript (default = 25). This means that if only a segment overlaps a transcript more than this threshold, the transcript is assigned the segment's SCNA event.

MCR_overlap_threshold	the percentage threshold for the overlap between a gene and an MCR region (default = 25). This means that if only a gene overlaps an MCR region more than this threshold, the gene is assigned the SCNA density of the MCR
hotspot_threshold	to determine hotspot genes, the (integer) threshold for the minimum number of cases with certain mutation in COSMIC (default = 5)
log2_hom_loss_threshold	to determine double-hit events, the \log_2 threshold for identifying homozygous loss events (default = -1).
verbose	boolean controlling verbosity (default = TRUE)
na.string	string that was used to indicate when a score is not available during annotation with ANNOVAR (default = '.')

Value

If prep_phenolyzer_input=FALSE (default), a data frame of features for prioritizing cancer driver genes (gene_symbol as the first column and 26 other columns containing features). If prep_phenolyzer_input=TRUE, the functions returns a vector gene symbols (union of all gene symbols for which scores are available) to be used as the input for performing 'phenolyzer' analysis.

The features data frame contains the following columns:

gene_symbol	HGNC gene symbol
metaprediction_score	the maximum metapredictor (coding) impact score for the gene
noncoding_score	the maximum non-coding PHRED-scaled CADD score for the gene
scna_score	SCNA proxy score. SCNA density (SCNA/Mb) of the minimal common region (MCR) in which the gene is located
hotspot_double_hit	boolean indicating whether the gene is a hotspot gene (indication of oncogenes) or subject to double-hit (indication of tumor-suppressor genes)
phenolyzer_score	'phenolyzer' score for the gene
hsa03320	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04010	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04020	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04024	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04060	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04066	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04110	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04115	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04150	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04151	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04210	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04310	boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04330 boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04340 boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04350 boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04370 boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04510 boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04512 boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04520 boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04630 boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04915 boolean indicating whether or not the gene takes part in this KEGG pathway

See Also

[prioritize_driver_genes](#) for prioritizing cancer driver genes

Examples

```
path2annovar_csv <- system.file('extdata/example.hg19_multianno.csv',  
                                package = 'driveR')  
path2phenolyzer_out <- system.file('extdata/example.annotated_gene_list',  
                                   package = 'driveR')  
features_df <- create_features_df(annovar_csv_path = path2annovar_csv,  
                                 scna_segs_df = example_scna_table,  
                                 phenolyzer_annotated_gene_list_path = path2phenolyzer_out)
```

create_gene_level_scna_df

Create Gene-level SCNA Data Frame

Description

Create Gene-level SCNA Data Frame

Usage

```
create_gene_level_scna_df(  
  scna_segs_df,  
  build = "GRCh37",  
  gene_overlap_threshold = 25  
)
```

Arguments

scna_segs_df the SCNA segments data frame. Must contain:

- chr** chromosome the segment is located in
- start** start position of the segment
- end** end position of the segment
- log2ratio** \log_2 ratio of the segment

build genome build for the SCNA segments data frame (default = 'GRCh37')

gene_overlap_threshold the percentage threshold for the overlap between a segment and a transcript (default = 25). This means that if only a segment overlaps a transcript more than this threshold, the transcript is assigned the segment's SCNA event.

Value

data frame of gene-level SCNA events, i.e. table of genes overlapped by SCNA segments.

create_noncoding_impact_score_df
Create Non-coding Impact Score Data Frame

Description

Create Non-coding Impact Score Data Frame

Usage

```
create_noncoding_impact_score_df(annovar_csv_path, na.string = ".")
```

Arguments

annovar_csv_path path to 'ANNOVAR' csv output file

na.string string that was used to indicate when a score is not available during annotation with ANNOVAR (default = '.')

Value

data frame of meta-prediction scores containing 2 columns:

- gene_symbol** HGNC gene symbol
- CADD_phred** PHRED-scaled CADD score

create_SCNA_score_df *Create SCNA Score Data Frame*

Description

Create SCNA Score Data Frame

Usage

```
create_SCNA_score_df(  
  scna_genes_df,  
  build = "GRCh37",  
  log2_ratio_threshold = 0.25,  
  MCR_overlap_threshold = 25  
)
```

Arguments

scna_genes_df data frame of gene-level SCNAs (can be output of [create_gene_level_scna_df](#))

build genome build for the SCNA segments data frame (default = 'GRCh37')

log2_ratio_threshold
the \log_2 ratio threshold for keeping high-confidence SCNA events (default = 0.25)

MCR_overlap_threshold
the percentage threshold for the overlap between a gene and an MCR region (default = 25). This means that if only a gene overlaps an MCR region more than this threshold, the gene is assigned the SCNA density of the MCR

Details

The function first aggregates SCNA \log_2 ratio on gene-level (by keeping the ratio with the maximal $|\log_2|$ ratio over all the SCNA segments overlapping a gene). Next, it identifies the minimal common regions (MCRs) that the genes overlap and finally assigns the SCNA density (SCNA/Mb) values as proxy SCNA scores.

Value

data frame of SCNA proxy scores containing 2 columns:

gene_symbol HGNC gene symbol

SCNA_density SCNA proxy score. SCNA density (SCNA/Mb) of the minimal common region (MCR) in which the gene is located.

`determine_double_hit_genes`*Determine Double-Hit Genes*

Description

Determine Double-Hit Genes

Usage

```
determine_double_hit_genes(  
  annovar_csv_path,  
  scna_genes_df,  
  log2_hom_loss_threshold = -1,  
  batch_analysis = FALSE  
)
```

Arguments

`annovar_csv_path` path to 'ANNOVAR' csv output file

`scna_genes_df` data frame of gene-level SCNAs (can be output of [create_gene_level_scna_df](#))

`log2_hom_loss_threshold` to determine double-hit events, the \log_2 threshold for identifying homozygous loss events (default = -1).

`batch_analysis` boolean to indicate whether to perform batch analysis (TRUE, default) or personalized analysis (FALSE). If TRUE, a column named 'tumor_id' should be present in both the ANNOVAR csv and the SCNA table.

Value

vector of gene symbols that are subject to double-hit event(s), i.e. non-synonymous mutation + homozygous copy-number loss

`determine_hotspot_genes`*Determine Hotspot Containing Genes*

Description

Determine Hotspot Containing Genes

Usage

```
determine_hotspot_genes(annovar_csv_path, hotspot_threshold = 5L)
```

Arguments

`annovar_csv_path`
path to 'ANNOVAR' csv output file

`hotspot_threshold`
to determine hotspot genes, the (integer) threshold for the minimum number of cases with certain mutation in COSMIC (default = 5)

Value

vector of gene symbols of genes containing hotspot mutation(s)

driveR	<i>driveR: An R Package for Prioritizing Cancer Driver Genes Using Genomics Data</i>
--------	--

Description

Cancer genomes contain large numbers of somatic alterations but few genes drive tumor development. Identifying cancer driver genes is critical for precision oncology. Most of current approaches either identify driver genes based on mutational recurrence or using estimated scores predicting the functional consequences of mutations.

Details

driveR is a tool for personalized or batch analysis of genomic data for driver gene prioritization by combining genomic information and prior biological knowledge. As features, driveR uses coding impact metaprediction scores, non-coding impact scores, somatic copy number alteration scores, hotspot gene/double-hit gene condition, 'phenolyzer' gene scores and memberships to cancer-related KEGG pathways. It uses these features to estimate cancer-type-specific probabilities for each gene of being a cancer driver using the related task of a multi-task learning classification model.

Author(s)

Maintainer: Ege Ulgen <egeulgen@gmail.com> ([ORCID](#)) [copyright holder]

See Also

[predict_coding_impact](#) for metaprediction of impact of coding variants. [create_features_df](#) for creating the features table to be used to prioritize cancer driver genes. See [prioritize_driver_genes](#) for prioritizing cancer driver genes

example_cohort_features_table

Example Cohort-level Features Table for Driver Prioritization

Description

The example dataset containing features for prioritizing cancer driver genes for 10 randomly selected samples from TCGA's LAML (Acute Myeloid Leukemia) cohort

Usage

example_cohort_features_table

Format

A data frame with 349 rows and 27 variables:

gene_symbol HGNC gene symbol

metaprediction_score the maximum metapredictor (coding) impact score for the gene

noncoding_score the maximum non-coding PHRED-scaled CADD score for the gene

scna_score SCNA proxy score. SCNA density (SCNA/Mb) of the minimal common region (MCR) in which the gene is located

hotspot_double_hit boolean indicating whether the gene is a hotspot gene (indication of oncogenes) or subject to double-hit (indication of tumor-suppressor genes)

phenolyzer_score 'phenolyzer' score for the gene

hsa03320 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04010 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04020 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04024 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04060 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04066 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04110 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04115 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04150 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04151 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04210 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04310 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04330 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04340 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04350 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04370 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04510 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04512 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04520 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04630 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04915 boolean indicating whether or not the gene takes part in this KEGG pathway

See Also

[KEGG_cancer_pathways_descriptions](#) for descriptions of KEGG 'Pathways in cancer'-related pathways.

example_cohort_scna_table

Example Cohort-level Somatic Copy Number Alteration Table

Description

A data set containing the somatic copy number alteration data for 10 randomly selected samples from TCGA's LAML (Acute Myeloid Leukemia) cohort

Usage

example_cohort_scna_table

Format

A data frame with 126147 rows and 5 variables:

chr chromosome the segment is located in

start start position of the segment

end end position of the segment

log2ratio \log_2 ratio of the segment

tumor_id ID for the tumor containing the SCNA segment

Source

https://dcc.icgc.org/releases/release_28

example_features_table

Example Features Table for Driver Prioritization

Description

The example dataset containing features for prioritizing cancer driver genes for the lung adenocarcinoma patient studied in Imielinski M, Greulich H, Kaplan B, et al. Oncogenic and sorafenib-sensitive ARAF mutations in lung adenocarcinoma. J Clin Invest. 2014;124(4):1582-6.

Usage

example_features_table

Format

A data frame with 4901 rows and 27 variables:

gene_symbol HGNC gene symbol

metaprediction_score the maximum metapredictor (coding) impact score for the gene

noncoding_score the maximum non-coding PHRED-scaled CADD score for the gene

scna_score SCNA proxy score. SCNA density (SCNA/Mb) of the minimal common region (MCR) in which the gene is located

hotspot_double_hit boolean indicating whether the gene is a hotspot gene (indication of oncogenes) or subject to double-hit (indication of tumor-suppressor genes)

phenolyzer_score 'phenolyzer' score for the gene

hsa03320 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04010 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04020 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04024 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04060 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04066 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04110 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04115 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04150 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04151 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04210 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04310 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04330 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04340 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04350 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04370 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04510 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04512 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04520 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04630 boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04915 boolean indicating whether or not the gene takes part in this KEGG pathway

See Also

[KEGG_cancer_pathways_descriptions](#) for descriptions of KEGG 'Pathways in cancer'-related pathways.

example_gene_scna_table

Example Gene-level Somatic Copy Number Alteration Table

Description

A data set containing the gene-level somatic copy number alteration data for the lung adenocarcinoma patient studied in Imielinski M, Greulich H, Kaplan B, et al. Oncogenic and sorafenib-sensitive ARAF mutations in lung adenocarcinoma. J Clin Invest. 2014;124(4):1582-6.

Usage

example_gene_scna_table

Format

A data frame with 46270 rows and 2 variables:

symbol gene symbol

log2ratio \log_2 ratio of the gene

Source

<https://pubmed.ncbi.nlm.nih.gov/24569458/>

example_scna_table *Example Somatic Copy Number Alteration Table*

Description

A data set containing the somatic copy number alteration data for the lung adenocarcinoma patient studied in Imielinski M, Greulich H, Kaplan B, et al. Oncogenic and sorafenib-sensitive ARAF mutations in lung adenocarcinoma. J Clin Invest. 2014;124(4):1582-6.

Usage

example_scna_table

Format

A data frame with 3160 rows and 4 variables:

chr chromosome the segment is located in

start start position of the segment

end end position of the segment

log2ratio \log_2 ratio of the segment

Source

<https://pubmed.ncbi.nlm.nih.gov/24569458/>

KEGG_cancer_pathways_descriptions

KEGG 'Pathways in cancer'-related Pathways - Descriptions

Description

A data frame containing descriptions for KEGG 'Pathways in cancer' (hsa05200)-related pathways. Generated on Nov 17, 2020.

Usage

KEGG_cancer_pathways_descriptions

Format

A data frame with 21 rows and 2 variables:

id KEGG pathway ID

description KEGG pathway description

 MTL_submodel_descriptions

MTL Sub-model Descriptions

Description

A data frame containing descriptions for all sub-models of the MTL model.

Usage

```
MTL_submodel_descriptions
```

Format

A data frame with 21 rows and 2 variables:

short_name short name for the cancer type

description description of the cancer type

```
predict_coding_impact Create Coding Impact Meta-prediction Score Data Frame
```

Description

Create Coding Impact Meta-prediction Score Data Frame

Usage

```
predict_coding_impact(
  annovar_csv_path,
  keep_highest_score = TRUE,
  keep_single_symbol = TRUE,
  na.string = "."
)
```

Arguments

```
annovar_csv_path
  path to 'ANNOVAR' csv output file
keep_highest_score
  boolean to indicate whether to keep only the maximal impact score per gene
  (default = TRUE). If FALSE, all scores per each gene are returned
keep_single_symbol
  in ANNOVAR outputs, a variant may be annotated as exonic in multiple genes.
  This boolean argument controls whether or not to keep only the first encountered
  symbol for a variant (default = TRUE)
na.string
  string that was used to indicate when a score is not available during annotation
  with ANNOVAR (default = '.')
```

Value

data frame of meta-prediction scores containing 2 columns:

gene_symbol HGNC gene symbol

metaprediction_score metapredictor impact score

Examples

```
path2annovar_csv <- system.file('extdata/example.hg19_multianno.csv',
                                package = 'driveR')
metapred_df <- predict_coding_impact(path2annovar_csv)
```

prioritize_driver_genes

Prioritize Cancer Driver Genes

Description

Prioritize Cancer Driver Genes

Usage

```
prioritize_driver_genes(features_df, cancer_type)
```

Arguments

features_df the features data frame for all genes, containing the following columns:

- gene_symbol** HGNC gene symbol
- metaprediction_score** the maximum metapredictor (coding) impact score for the gene
- noncoding_score** the maximum non-coding PHRED-scaled CADD score for the gene
- scna_score** SCNA proxy score. SCNA density (SCNA/Mb) of the minimal common region (MCR) in which the gene is located
- hotspot_double_hit** boolean indicating whether the gene is a hotspot gene (indication of oncogenes) or subject to double-hit (indication of tumor-suppressor genes)
- phenolyzer_score** 'phenolyzer' score for the gene
- hsa03320** boolean indicating whether or not the gene takes part in this KEGG pathway
- hsa04010** boolean indicating whether or not the gene takes part in this KEGG pathway
- hsa04020** boolean indicating whether or not the gene takes part in this KEGG pathway
- hsa04024** boolean indicating whether or not the gene takes part in this KEGG pathway

hsa04060	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04066	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04110	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04115	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04150	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04151	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04210	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04310	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04330	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04340	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04350	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04370	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04510	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04512	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04520	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04630	boolean indicating whether or not the gene takes part in this KEGG pathway
hsa04915	boolean indicating whether or not the gene takes part in this KEGG pathway
cancer_type	short name of the cancer type. All available cancer types are listed in MTL_submodel_descriptions

Value

data frame with 3 columns:

gene_symbol HGNC gene symbol

driverness_prob estimated probability for each gene in features_df of being a cancer driver. The probabilities are calculated using the selected (via cancer_type) cancer type's sub-model.

prediction prediction based on the cancer-type-specific threshold (either 'driver' or 'non-driver')

See Also

[create_features_df](#) for creating the features table.

Examples

```
drivers_df <- prioritize_driver_genes(example_features_table, 'LUAD')
```

specific_thresholds *Tumor type specific probability thresholds*

Description

Driver gene probability thresholds for all 21 cancer types (submodels).

Usage

```
specific_thresholds
```

Format

vector with 21 elements

Index

* datasets

- example_cohort_features_table, [10](#)
- example_cohort_scna_table, [11](#)
- example_features_table, [12](#)
- example_gene_scna_table, [13](#)
- example_scna_table, [14](#)
- KEGG_cancer_pathways_descriptions, [14](#)
- MTL_submodel_descriptions, [15](#)
- specific_thresholds, [18](#)

- create_features_df, [2](#), [9](#), [18](#)
- create_gene_level_scna_df, [3](#), [5](#), [7](#), [8](#)
- create_noncoding_impact_score_df, [6](#)
- create_SCNA_score_df, [7](#)

- determine_double_hit_genes, [8](#)
- determine_hotspot_genes, [8](#)
- driveR, [9](#)
- driveR-package (driveR), [9](#)

- example_cohort_features_table, [10](#)
- example_cohort_scna_table, [11](#)
- example_features_table, [12](#)
- example_gene_scna_table, [13](#)
- example_scna_table, [14](#)

- KEGG_cancer_pathways_descriptions, [11](#), [13](#), [14](#)

- MTL_submodel_descriptions, [15](#), [17](#)

- predict_coding_impact, [9](#), [15](#)
- prioritize_driver_genes, [5](#), [9](#), [16](#)

- specific_thresholds, [18](#)