Package ‘depmap’

October 12, 2020

Type Package

Title Cancer Dependency Map Data Package

Version 1.2.0

Description The depmap package is a data package that accesses datsets from the Broad Institute DepMap cancer dependency study using ExperimentHub. Datasets from the most current release are available, including RNAI and CRISPR-Cas9 gene knockout screens quantifying the genetic dependency for select cancer cell lines. Additional datasets are also available pertaining to the log copy number of genes for select cell lines, protein expression of cell lines as measured by reverse phase protein lysate microarray (RPPA), 'Transcript Per Million' (TPM) data, as well as supplementary datasets which contain metadata and mutation calls for the other datasets found in the current release. The 19Q3 release adds the drug_dependency dataset, that contains cancer cell line dependency data with respect to drug and drug-candidate compounds. This package will be updated on a quarterly basis to incorporate the latest Broad Institute DepMap Public cancer dependency datasets. All data made available in this package was generated by the Broad Institute DepMap for research purposes and not intended for clinical use. This data is distributed under the Creative Commons license (Attribution 4.0 International (CC BY 4.0)).

Depends R (>= 3.6), methods, dplyr

Imports utils, ExperimentHub, AnnotationHub

License Artistic-2.0

Encoding UTF-8

RoxygenNote 7.1.0

Suggests knitr, rmarkdown, BiocStyle, viridis, gridExtra, ggplot2, readr, tibble, stringr, tidyr

VignetteBuilder knitr

biocViews ExperimentHub, ExperimentData, ReproducibleResearch, RepositoryData, AssayDomainData, CopyNumberVariationData, DiseaseModel, CancerData, BreastCancerData, ColonCancerData, KidneyCancerData, LeukemiaCancerData, LungCancerData, OvarianCancerData, ProstateCancerData, OrganismData, Homo_sapiens_Data, PackageTypeData, SpecimenSource, CellCulture, Genome, Proteome, StemCell, Tissue

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

git_branch RELEASE_3_11

git_last_commit de96217

git_last_commit_date 2020-04-27

Date/Publication 2020-10-12

Author Laurent Gatto [cre, aut],
Theo Killian [aut]

Maintainer Laurent Gatto <`laurent.gatto@uclouvain.be`>

R topics documented:

`copyNumber` *copyNumber_20Q1*

Description

The ‘copyNumber‘ dataset contains the 20Q1 WES copy number data for genes and cancer cell lines. This dataset includes data from 27639 genes, 1713 cell lines, 35 primary diseases and 36 lineages. The columns of ‘copyNumber‘ are: ‘depmap_id‘, a foreign key corresponding to the cancer cell lineage, ‘cell_line‘ containing the common CCLE name of the cancer cell lines, ‘gene‘ containing both the HUGO gene name of the knockdown gene along with entrez ID#, ‘gene_name‘ containing only the HUGO gene name, ‘entrez_id‘ containing only the entrez ID#, and ‘log_copy_number‘ containing the numerical dependency score values for each pair of genes and cell lines. This dataset can be loaded into the R environment with the ‘depmap_copyNumber‘ function.

Usage

```
copyNumber
```

Format

A data frame with 47345607 rows (cell lines) and 6 variables.

depmap_id Cancer cell line foreign key (i.e. "ACH-00001")

gene HUGO symbol (e.g. "SAP25") and Entrez ID# (e.g. 100316904)

gene_name HUGO symbol (e.g. "SAP25")

entrez_id Entrez ID# (e.g. 100316904)

log_copy_number numerical log fold change in copy number for a given gene and cell line

cell_line CCLE name of cancer cell line (i.e. "184A1_BREAST")

Details

This data represents the ‘CCLE_gene_cn.csv‘ file taken from the 20Q1 [Broad Institute](https://depmap.org/portal/downl
cancer depenedency study. The derived dataset found in the ‘depmap‘ package features the addition of a foreign key ‘depmap_id‘ found in the first column of this dataset, which was added from the ‘metadata‘ dataset. This dataset has been converted to a long format tibble. Variables names from the original dataset were converted to lower case, put in snake case, and abbreviated where feasible.

Change log

- 19Q1: Initial dataset consisted of a data frame with 37907473 rows (cell lines) and 6 variables representing 23299 genes, 1604 cell lines, 38 primary diseases and 33 lineages.

- 19Q2: adds 23 cell lines and 1 lineage

- 19Q3: adds 3263 genes, 30 cell lines and removes 2 primary diseases. Now a dataframe with 45670234 rows and 6 variables.

- 19Q4: adds 77 genes, 25 cell lines, 0 primary diseases and 2 lineages.

- 20Q1: adds 31 cell lines

Source

DepMap, Broad (2019)

References

Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., ... & Meyers, R. M. (2017). Defining a cancer dependency map. Cell, 170(3), 564-576. (PubMed)

DepMap, Broad (2019): DepMap Achilles 19Q1 Public. (figshare). Fileset.

Robin M. Meyers, Jordan G. Bryan, James M. McFarland, Barbara A. Weir, ... David E. Root, William C. Hahn, Aviad Tsherniak. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nature Genetics 2017 October 49:1779–1784. (Pubmed)

Mahmoud Ghandi, Franklin W. Huang, Judit Jané-Valbuena, Gregory V. Kryukov, ... Todd R. Golub, Levi A. Garraway & William R. Sellers. 2019. Next- generation characterization of the Cancer Cell Line Encyclopedia. Nature 569, 503–508 (2019). (Nature)

Examples

```
## Not run:
depmap_copyNumber()

## End(Not run)
```

`crispr` *crispr_20Q1*

Description

The ‘crispr‘ dataset contains the 20Q1 genetic dependency of CRISPR-Cas9 gene knockout of select genes in various cancer cell lines. This dataset includes data from 18333 genes, 739 cell lines, 29 primary diseases and 26 lineages. The columns of ‘crispr‘ are: ‘depmap_id‘ a foreign key corresponding to the cancer cell lineage, ‘cell_line‘ containing the common CCLE name of the cancer cell lines, ‘gene‘ containing both the HUGO gene name of the knockout gene along with entrez ID#, ‘gene_name‘ containing only the HUGO gene name, ‘entrez_id‘ containing only the entrez ID# and ‘dependency‘ which contains the numerical dependency score values for each pair of genes and cell lines. This dataset can be loaded into R environment with the ‘depmap_crispr‘ function.

Usage

```
crispr
```

Format

A data frame with 13548087 rows (cell lines) and 6 variables:

depmap_id Cancer cell line foreign key (i.e. "ACH-00001")

gene HUGO symbol (e.g. "SAP25") and Entrez ID# (e.g. 100316904)

gene_name HUGO symbol (e.g. "SAP25")

entrez_id Entrez ID# (e.g. 100316904)

dependency numerical depenency score of given gene and cell line

cell_line CCLE name of cancer cell line (i.e. "184A1_BREAST")

Details

This data represents the ‘Achilles_gene_effect.csv‘ file taken from the 20Q1 [Broad Institute](https://depmap.org/portal/d cancer depenedency study. The derived dataset found in the ‘depmap‘ package features the addition of a foreign key ‘depmap_id‘ found in the first column of this dataset, which was added from the ‘metadata‘ dataset. This dataset has been converted to a long format tibble. Variables names from the original dataset were converted to lower case, put in snake case, and abbreviated where feasible.

Change log

- 19Q1: Initial dataset consisted of a data frame with 9839772 rows (cell lines) and 6 variables representing 17634 genes, 558 cell lines, 26 primary diseases and 28 lineages.

- 19Q2: adds 5 cell lines, 1 primary disease and 1 lineage

- 19Q3: adds 699 genes, 62 cell lines and 1 primary disease. Now a dataset with 11458125 rows and 6 variables.

- 19Q4: adds 0 genes, 64 cell lines, 1 primary diseases and 0 lineages.

- 20Q1: adds 50 cell lines

Source

DepMap, Broad (2019)

References

Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., ... & Meyers, R. M. (2017). Defining a cancer dependency map. Cell, 170(3), 564-576. (PubMed)

DepMap, Broad (2019): DepMap Achilles 19Q1 Public. (figshare). Fileset.

Robin M. Meyers, Jordan G. Bryan, James M. McFarland, Barbara A. Weir, ... David E. Root, William C. Hahn, Aviad Tsherniak. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nature Genetics 2017 October 49:1779–1784. (Pubmed)

Mahmoud Ghandi, Franklin W. Huang, Judit Jané-Valbuena, Gregory V. Kryukov, ... Todd R. Golub, Levi A. Garraway & William R. Sellers. 2019. Next- generation characterization of the Cancer Cell Line Encyclopedia. Nature 569, 503–508 (2019). (Nature)

Examples

```
## Not run:
depmap_crispr()

## End(Not run)
```

`depmap` *depmap: Cancer Dependency Map Data Package*

Description

The depmap package is a data package that accesses datsets from the Broad Institute DepMap cancer dependency study using ExperimentHub. Datasets from the most current release are available, including RNAI and CRISPR-Cas9 gene knockout screens quantifying the genetic dependency for select cancer cell lines. Additional datasets are also available pertaining to the log copy number of genes for select cell lines, protein expression of cell lines as measured by reverse phase protein lysate microarray (RPPA), 'Transcript Per Million' (TPM) data, chemical dependecy (drug_sensativity) as well as supplementary datasets which contain metadata and mutation calls for the other datasets found in the current release. This package will be updated on a quarterly basis to incorporate the latest Broad Institute DepMap Public cancer dependency datasets. All data made available in this package was generated by the Broad Institute DepMap for research purposes and not intended for clinical use. This data is distributed under the Creative Commons license (Attribution 4.0 International (CC BY 4.0)).

Details

See the package vignettes and respective man pages for details.

`depmap_release` *Returns the depmap release*

Description

Returns the depmap release

Usage

```
depmap_release()
```

Value

`character(1)` with the depmap release

`drug_sensitivity` *drug_sensitivity_19Q3*

Description

The ‘drug_sensitivity‘ dataset contains the 19Q3 replicate collapsed logfold change values relative to DMSO, corrected for experimental confounders using ComBat. This dataset contains information refering to 4686 compounds, 578 cell lines, 23 primary diseases and 25 lineages. This dataset is part of the SIGMA Repurposing release which contains small molecule viability datasets generated using the Broad Repurposing Library and the PRISM multiplexed cell-line viability assay. The columns of ‘drug_sensitivity‘ are: ‘depmap_id‘ a foreign key corresponding to the cancer cell lineage, ‘cell_line‘ the common CCLE name of the cancer cell lines, ‘compound‘ the synonym for the drug compound, and ‘dependency‘ which contains the numerical dependency score values for each pair of genes and cell lines.

Usage

```
drug_sensitivity
```

Format

A data frame with 67498602 rows (cell lines) and 6 variables:

depmap_id Cell line foreign key (i.e. "ACH-000956")

cell_line Name of cancer cell line (i.e. "22RV1_PROSTATE")

compound Drug compound name (i.e. BRD-A00077618-236-07-6::2.5::HTS)

dependency numerical depenency score of a gene for a cell line

Details

This data originates from the ‘primary_replicate_collapsed_logfold_change.csv‘ file taken from the 19Q3 [Broad Institute](https://depmap.org/portal/download/) cancer depenedency study. The derived dataset found in the ‘depmap‘ package features the addition of a foreign key ‘depmap_id‘ found in the first column of this dataset, which was added from the ‘metadata‘ dataset. This dataset has been converted to a long format tibble. Variables names from the original dataset were converted to lower case, put in snake case, and abbreviated where feasible.

Change log

- 19Q3: Initial dataset consisted of a data frame with 2708508 rows (cell lines) and 6 variables representing 686 compounds, 578 cell lines, 23 primary diseases and 25 lineages.

- 19Q4: no change, no further releases are scheduled at this time.

- 20Q1: no change, no further releases are scheduled at this time.

Source

DepMap, Broad (2019)

References

Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., ... & Meyers, R. M. (2017). Defining a cancer dependency map. Cell, 170(3), 564-576. (PubMed)

Steven M Corsello, Rohith T Nagari, Ryan D Spangler, Jordan Rossen, Mustafa Kocak, Jordan G Bryan, Ranad Humeidi, David Peck, Xiaoyun Wu, Andrew A Tang, Vickie MWang, Samantha A Bender, Evan Lemire, Rajiv Narayan, Philip Montgomery, Uri Ben-David, Yejia Chen, Matthew G Rees, Nicholas J Lyons, James M McFarland, Bang TWong, Li Wang, Nancy Dumont, Patrick J O'Hearn, Eric Stefan, John G Doench, HeidiGreulich, Matthew Meyerson, Francisca Vazquez, Aravind Subramanian, Jennifer A Roth, Joshua A Bittker, Jesse S Boehm, Christopher C Mader, Aviad Tsherniak, Todd R Golub. 2019. Non-oncology drugs are a source of previously unappreciated anti-cancer activity. (bioRXiv)

`metadata` *metadata_20Q1*

Description

The ‘metadata‘ dataset contains the metadata about cell lines in the 20Q1 Broad Institute DepMap release, which includes mapping between ‘depmap_id‘ and ‘cell_line‘ name for cancer cell lines. This dataset does not contain any dependency data but contains the metadata for 0 genes, 1775 cell lines, 35 primary diseases and 37 lineages. The columns of ‘metadata‘ are: ‘depmap_id‘, ‘stripped_cell_line_name‘, ‘cell_line‘, ‘aliases‘, ‘cosmic_id‘, ‘sanger_id‘, ‘primary_disease‘, ‘subtype_disease‘, ‘sub_subtype_disease‘, ‘gender‘, ‘source‘, ‘Achilles_n_replicates‘, ‘cell_line_NNMD‘, ‘culture_type‘, ‘culture_medium‘, and ‘cas9_activity‘. This dataset can be loaded into the R environment with the ‘depmap_copyNumber‘ function.

Usage

```
metadata
```

Format

A data frame with 1755 rows (cell lines) and 16 variables:

depmap_id Cancer cell line primary key, used in other datasets as foreign key (i.e. "ACH-00001")

stripped_cell_line_name Name of stripped cell line

cell_line CCLE name of cancer cell line (i.e. "184A1_BREAST")

aliases Aliases of cancer cell line

cosmic_id Catalogue Of Somatic Mutations In Cancer (COSMIC) ID number (e.g. 905933)

lineage Lineage of cancer cell line

lineage_subtype Subtype of lineage of cancer cell line

lineage_sub_subtype Subtype of subtype of Lineage of cancer cell line

lineage_molecular_subtype Molecular type of Lineage of cancer cell line

sex Sex of tissue sample)

source Source of tissue sample)

Achilles_n_replicates Number of replicates)

cell_line_NNMD Cell line NNMD)

culture_type Culture type of tissue sample)

culture_medium Culture medium of tissue sample)

cas9_activity Cas9 activity)

RRID Resource Identification Portal ID

sample_collection_site Site of sample collection

primary_or_metastasis Primary cancer cell line or metastatic

primary_disease Primary Disease (e.g. cancer type)

subtype_disease Subtype Disease (e.g. Acute Myelogenous Leukemia (AML), M3 (Promyelocytic))

age Age of individual sample of cell line was derived

sanger_id Sanger ID (eg. 2201)

additional_info Additional information about samples

Details

This data represents the ‘sample_info.csv‘ file taken from the 20Q1 [Broad Institute](https://depmap.org/portal/download cancer depenedency study. This dataset features the a primary key ‘depmap_id‘ which is a unique ID given to each cell line and is found in the first column of this dataset. The ‘depmap_id‘ attribute is used as a foreign key in all other datasets in the package. This dataset has been converted to a long format tibble. This dataset does not contain any expression or dependency data but rather contains the metadata for all cancer cell lines used in the depmap project. Variables names were converted to lower case, put in snake case, and abbreviated where feasible (e.g. "Sanger ID" was changed to "sanger_id").

Change log

- 19Q1: Initial dataset consisted of data frame with 1677 rows (cell lines) and 9 variables, representing 0 genes, 1677 cell lines, 38 primary diseases and 33 lineages

- 19Q2: adds 37 new cell lines, 1 primary disease and 1 lineage. This version of the metadata dataset contains 6 variables not found in previous versions, including: ‘Achilles_n_replicates‘, ‘cell_line_NNMD‘, ‘culture_type‘, ‘culture_medium‘, and ‘cas9_activity‘.

- 19Q3: adds 0 genes, 30 cell lines, 2 primary diseases and 2 lineages.

- 19Q4: for 0 genes, 42 cell lines, 0 primary diseases and 3 lineages.

- 20Q1: adds 19 cell lines, ‘gender‘ was changed to ‘sex‘, ‘age‘, ‘primary_or_metastasis‘ and ‘sample_collection_site“ were added

Source

DepMap, Broad (2019)

References

Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., ... & Meyers, R. M. (2017). Defining a cancer dependency map. Cell, 170(3), 564-576. (PubMed)

DepMap, Broad (2019): DepMap Achilles 19Q1 Public. (figshare). Fileset.

Robin M. Meyers, Jordan G. Bryan, James M. McFarland, Barbara A. Weir, ... David E. Root, William C. Hahn, Aviad Tsherniak. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nature Genetics 2017 October 49:1779–1784. (Pubmed)14

Mahmoud Ghandi, Franklin W. Huang, Judit Jané-Valbuena, Gregory V. Kryukov, ... Todd R. Golub, Levi A. Garraway & William R. Sellers. 2019. Next- generation characterization of the Cancer Cell Line Encyclopedia. Nature 569, 503–508 (2019). (Nature)

Examples

```
## Not run:
depmap_metadata()

## End(Not run)
```

`mutationCalls` *mutationCalls_20Q1*

Description

The ‘mutationCalls‘ dataset contains merged the 20Q1 mutation calls (for coding region, germline filtered) and includes data from 18802 genes, 1697 cell lines, 35 primary diseases and 36 lineages. This dataset can be considered the metadata data set for mutations and does not contain any dependency data. This dataset can be loaded into the R environment with the ‘depmap_mutationCalls‘ function.

Usage

```
mutationCalls
```

Format

A data frame with 1279923 rows and 34 variables:

depmap_id depmap_id

gene_name Hugo Symbol denotes a unique and meaningful name for each gene (e.g. SAP25)

entrez_id Gene ID for NCBI Entrez gene database, (e.g. 100316904)

ncbi_build NCBI Build (i.e. reference genome)

chromosome Chromosome

start_pos Gene start position

end_pos Gene end position

strand Strand location of gene

var_class Variant Classification

var_type Variant Type

ref_allele Reference Allele

tumor_seq_allele1 Tumor Seq Allele1

dbSNP_RS Single Nucleotide Polymorphism Database (dbSNP)3 reference cluster

dbSNP_val_status dbSNP Val Status

genome_change Genome Change

annotation_transcript Annotation Transcript

tumor_sample_barcode Tumor Sample Barcode

cDNA_change change in cDNA

codon_change Codon_Change

protein_change Protein_Change

is_deleterious Status of gene knockout on cell lineage

is_tcga_hotspot isTCGAhotspot

tcga_hsCnt TCGAhsCnt

is_cosmic_hotspot isCOSMIChotspot

cosmic_hsCnt COSMIChsCnt

ExAC_AF ExAC_AF

CGA_WES_AC CGA_WES_AC

sanger_WES_AC SangerWES_AC

sanger_recalib_WES_AC SangerRecalibWES_AC

RNAseq_AC RNAseq_AC

HC_AC HC_AC

RD_AC RD_AC

WGS_AC WGS_AC

var_annotation Variant_annotation

Details

This data represents the ‘CCLE_mutations.csv‘ file taken from the 20Q1 [Broad Institute](https://depmap.org/portal/dow cancer depenedency study. The derived dataset found in the ‘depmap‘ package features the addition of a foreign key ‘depmap_id‘ found in the first column of this dataset, which was added from the ‘metadata‘ dataset. This dataset has been converted to a long format tibble. Variables names from the original dataset were converted to lower case, put in snake case, and abbreviated where feasible.

Change log

- 19Q1: Initial dataset for package consisted of dataframe with 1243145 rows and 35 variables representing 18755 genes, 1601 cell lines, 37 primary diseases and 33 lineages

- 19Q2: adds 30 cell lines, 1 primary disease and 1 lineage. This version has different columns than the previous version: the variable "VA_WES_AC" is no longer present in this dataset. Some minor alterations to the original file were made. The first column of the original dataset, (itemIDSample number) was removed, as this column was only the row number and did not serve any unique identifying purpose.

- 19Q3: adds 1 gene, 25 cell lines and removes 1 primary disease.

- 19Q4: adds 1 gene, 10 cell lines, 0 primary diseases and 2 lineages.

- 20Q1: adds 4 genes, 31 cell lines

Source

DepMap, Broad (2019)

References

Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., ... & Meyers, R. M. (2017). Defining a cancer dependency map. Cell, 170(3), 564-576. (PubMed)

DepMap, Broad (2019): DepMap Achilles 19Q1 Public. (figshare). Fileset.

Robin M. Meyers, Jordan G. Bryan, James M. McFarland, Barbara A. Weir, ... David E. Root, William C. Hahn, Aviad Tsherniak. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nature Genetics 2017 October 49:1779–1784. (Pubmed)

Mahmoud Ghandi, Franklin W. Huang, Judit Jané-Valbuena, Gregory V. Kryukov, ... Todd R. Golub, Levi A. Garraway & William R. Sellers. 2019. Next- generation characterization of the Cancer Cell Line Encyclopedia. Nature 569, 503–508 (2019). (Nature)

Examples

```
## Not run:
depmap_mutationCalls()

## End(Not run)
```

`rnai` *rnai_19Q3*

Description

The ‘rnai‘ dataset contains the 19Q3 cancer dependency of select cancer cell lines for genes found by RNAi gene knockdown. This dataset includes data from 17309 genes, 712 cancer cell lines, 31 primary diseases and 31 lineages. The columns of ‘rnai‘ are: ‘depmap_id‘, a foreign key corresponding to the cancer cell lineage, ‘cell_line‘ containing the common CCLE name of the cancer cell lines, ‘gene‘ containing both the HUGO gene name of the knockdown gene along with entrez ID#, ‘gene_name‘ which only contains HUGO gene name, ‘entrez_id‘ which contains only the entrez ID# and ‘dependency‘ which contains the numerical dependency score values for each pair of genes and cell lines. This dataset can be loaded into the R environment with the ‘depmap_rnai‘ function.

Usage

```
rnai
```

Format

A data frame with 12324008 rows (cell lines) and 6 variables:

depmap_id	cancer cancer cell line foreign key (i.e. "ACH-00001")
cell_line	CCLE name of cancer cell line (i.e. "184A1_BREAST")
gene	HUGO symbol (e.g. "SAP25") and Entrez ID# (e.g. 100316904)
gene_name	HUGO symbol (e.g. "SAP25")
entrez_id	Entrez ID# (e.g. 100316904)
dependency	numerical depenency score of a gene for a cell line

Details

This data represents the ‘D2_combined_genetic_dependency_scores‘ file taken from the 19Q3 [Broad Institute](https://depmap.org/portal/download/) cancer depenedency study. The derived dataset found in the ‘depmap‘ package features the addition of a foreign key ‘depmap_id‘ found in the first column of this dataset, which was added from the ‘metadata‘ dataset. This dataset has been converted to a long format tibble. Variables names from the original dataset were converted to lower case, put in snake case, and abbreviated where feasible.

Change log

- 19Q1: Initial dataset consisted of a data frame with 12324008 rows (cell lines) and 6 variables representing 17309 genes, 711 cancer cell lines, 30 primary diseases and 31 lineages.

- 19Q2: adds 1 cell line

- 19Q3: adds 1 primary disease

- 19Q4: no change, no further releases are scheduled at this time.

- 20Q1: no change, no further releases are scheduled at this time.

Source

DepMap, Broad (2019)

References

Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., ... & Meyers, R. M. (2017). Defining a cancer dependency map. Cell, 170(3), 564-576. (PubMed)

James M. McFarland, Zandra V. Ho, Guillaume Kugener, Joshua M. Dempster, Phillip G. Montgomery, Jordan G. Bryan, John M. Krill-Burger, Thomas M. Green, Francisca Vazquez, Jesse S. Boehm, Todd R. Golub, William C. Hahn, David E. Root, Aviad Tsherniak. (2018). Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. Nature Communications 9, 1. (Nature)

Examples

```
## Not run:
depmap_rnai()

## End(Not run)
```

`RPPA` *RPPA_19Q3*

Description

The ‘RPPA‘ dataset contains the 19Q3 CCLE Reverse Phase Protein Array (RPPA) cellular model expression data. This dataset includes data from 214 genes, 899 cancer cell lines, 28 primary diseases and 28 lineages. The columns of ‘RPPA‘ are: ‘depmap_id‘, a foreign key corresponding to the cancer cell lineage, ‘cell_line‘ which contains the common CCLE name of the cancer cell lines, ‘gene‘ which contains the knockdown gene expression, ‘antibody‘ containing the name of knocked down gene and ‘expression‘ containing numerical protein expression data. This dataset can be loaded into R environment with the ‘depmap_RPPA‘ function.

Usage

```
RPPA
```

Format

A data frame with 192386 rows and 4 variables:

depmap_id cancer cell line foreign key (i.e. "ACH-000001")

cell_line CCLE name of cancer cell line (i.e. "NIHOVCAR3_OVARY")

antibody Name of antibody targeting protein (i.e. "14-3-3_beta")

expression Observed expression via RPPA of protein coding genes

Details

This data represents the ‘CCLE_RPPA_20181003.csv‘ file taken from the 19Q3 [Broad Institute](https://depmap.org/port cancer depenedency study. The derived dataset found in the ‘depmap‘ package features the addition of a foreign key ‘depmap_id‘ found in the first column of this dataset, which was added from the ‘metadata‘ dataset. This dataset has been converted to a long format tibble. Variables names from the original dataset were converted to lower case, put in snake case, and abbreviated where feasible.

Change log

- 19Q1: Initial dataset consisted of a data frame with 192386 rows and 4 variables representing 214 genes, 899 cancer cell lines, 28 primary diseases and 28 lineages.

- 19Q2: no change

- 19Q3: no change

- 19Q4: no change, no further releases are scheduled at this time.

- 20Q1: no change, no further releases are scheduled at this time.

Source

DepMap, Broad (2019)

References

Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., ... & Meyers, R. M. (2017). Defining a cancer dependency map. Cell, 170(3), 564-576. (PubMed)

Mahmoud Ghandi, Franklin W. Huang, Judit Jané-Valbuena, Gregory V. Kryukov, ... Todd R. Golub, Levi A. Garraway & William R. Sellers. 2019. Next- generation characterization of the Cancer Cell Line Encyclopedia. Nature 569, 503–508 (2019). (Nature)

Haoxin Li, Shaoyang Ning, Mahmoud Ghandi, Gregory V. Kryukov, Shuba Gopal, ... Levi A. Garraway & William R. Sellers. The landscape of cancer cell line metabolism. Nature Medicine 25, 850-860 (2019). (Nature)

Examples

```
## Not run:
depmap_RPPA()

## End(Not run)
```

`TPM` *TPM_20Q1*

Description

The ‘TPM‘ dataset contains the 20Q1 CCLE "Transcript Per Million" RNAseq gene expression data for protein coding genes. This dataset includes data from 19144 genes, 1270 cell lines, 32 primary diseases and 34 lineages. The columns of ‘TPM‘ are: ‘depmap_id‘, a foreign key corresponding to the cancer cell lineage, ‘cell_line‘ the common CCLE name of the cancer cell lines, ‘gene‘ containing both the HUGO gene name of the knockdown gene along with ensembl ID#, ‘gene_name‘ containing the HUGO gene name and ‘ensembl_id‘ containing only the ensembl ID# and ‘expression‘ which contains the numerical protein coding gene expression change at scale (log2(TPM+1)). This dataset can be loaded into R environment with the ‘depmap_RPPA‘ function.

Usage

```
TPM
```

Format

A data frame with 24312880 rows (cell lines) and 6 variables:

depmap_id Cell line foreign key (i.e. "ACH-000956")

cell_line Name of cancer cell line (i.e. "22RV1_PROSTATE")

gene HUGO symbol and Ensembl ID (e.g. TSPAN6 (ENSG00000000003))

gene_name HUGO symbol (e.g. "TSPAN6")

ensembl_id Ensembl ID (e.g. ENSG00000044574)

expression Log fold (log2(TPM+1)) protein expression change

Details

This data originates from the ‘CCLE_expression.csv‘ file taken from the 20Q1 [Broad Institute](https://depmap.org/porta cancer depenedency study. The derived dataset found in the ‘depmap‘ package features the addition of a foreign key ‘depmap_id‘ found in the first column of this dataset, which was added from the ‘metadata‘ dataset. This dataset has been converted to a long format tibble. Variables names from the original dataset were converted to lower case, put in snake case, and abbreviated where feasible.

Change log

- 19Q1: Initial dataset consisted of a data frame with 67360300 rows (cell lines) and 6 variables representing 57820 genes, 1165 cell lines, 33 primary diseases, 32 lineages.

- 19Q2: removes 1618 genes, adds 36 cell lines, removes one primary disease and adds 1 lineage

- 19Q3: removes 37058 genes, adds 9 cell lines, removes 2 primary diseases. Now a 23164240 by 6 dataframe.

- 19Q4: 0 genes, 39 cancer cell lines, 0 primary diseases and 1 lineage

- 20Q1: adds 31 cell lines

Source

DepMap, Broad (2019)

References

Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., ... & Meyers, R. M. (2017). Defining a cancer dependency map. Cell, 170(3), 564-576. (PubMed)

DepMap, Broad (2019): DepMap Achilles 19Q1 Public. (figshare). Fileset.

Robin M. Meyers, Jordan G. Bryan, James M. McFarland, Barbara A. Weir, ... David E. Root, William C. Hahn, Aviad Tsherniak. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nature Genetics 2017 October 49:1779–1784. (Pubmed)

Mahmoud Ghandi, Franklin W. Huang, Judit Jané-Valbuena, Gregory V. Kryukov, ... Todd R. Golub, Levi A. Garraway & William R. Sellers. 2019. Next- generation characterization of the Cancer Cell Line Encyclopedia. Nature 569, 503–508 (2019). (Nature)

Examples

```
## Not run:
depmap_TPM()

## End(Not run)
```

Index