HT29benchmark R package

August 11, 2022

`HT29R.expNames` *Benchmark Experiment Names*

Description

Labels of individual HT-29 screen experiments

Usage

```
data("HT29R.expNames")
```

Format

A vector of 6 strings, containing each the name of one experiment.

Examples

```
data(HT29R.expNames)
print(HT29R.expNames)
```

`HT29R.downloadRefData` *Download reference HT-29 screens data.*

Description

This function allows downloading reference datasets from high-quality CRISPR-Cas9 pooled screens of the HT-29 cell line performed with the KY sgRNA library[1]. This data has been generated through the experimental pipeline described in [2] and it is also public aviable on the Project Score web-site (https://score.depmap.sanger.ac.uk/downloads), part of the Cancer Dependency Map portfolio of tools and resources at the Wellcome Sanger Institute (https://depmap.sanger.ac.uk/).

Usage

```
H29R.downloadRefData(
      whatToDownload = "FCs",
      destFolder = "./",
    dataRepoURL = "https://cog.sanger.ac.uk/cmp/downloads/crispr_cas9_benchmark/",
      expNames=c("HT29_c903","HT29_c904","HT29_c905",
                 "HT29_c906","HT29_c907","HT29_c908"))
```

Arguments

whatToDownload	String parameter specifying what type of data to dowload. Possible values are "rawCounts" for plain .tsv files containing raw sgRNA countsare or "FCs" (default) for R objects containing sgRNA normalised depletion fold-changes: data frames in which the first two columns contain sgRNAs' identifiers and HGNC symbols of targeted gene, followed by one column per screen replicate containing sgRNAs' fold-changes;
destFolder	String specifying where the dataset should be saved;
dataRepoURL	The URL of the data repository;
expNames	A vector of strings specifying the experiment names for the dataset to download.

Author(s)

Ichcha Manipur & Francesco Iorio (fi1@sanger.ac.uk)

References

[1] Tzelepis K, Koike-Yusa H, De Braekeleer E, Li Y, Metzakopian E, Dovey OM, et al. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. Cell Rep. 2016;17:1193–205.

[2] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

Examples

```
#### creating a temporary directory
dir.create('tempDir')

#### downloading reference sgRNA depletion fold-changes from high-quality
#### HT-29 screens into the temporary directory
HT29R.downloadRefDataset(destFolder = 'tempDir')
```

HT29R.FCdistributions *Screens depletion fold-changes (FCs) average statistics and distributions.*

Description

Average parameters (and confidence intervals) of sgRNA logFCs distributions observed when screening HT-29 with reagent and experimental settings described in Behan et al., plus user-provided screen statistics.

Usage

```
HT29R.FCdistributions(refDataDir='./',
                      resDir='./',
                      userFCs=NULL,
                      stats=TRUE,
                      saveToFig=TRUE,
                      display=FALSE)
```

Arguments

refDataDir	Reference HT29 dataset directory: a string specifying the location of the processed HT29 reference dataset.
resDir	Output directory: a string specifying the directory where the output of this function (a pdf file with multiple plots) should be saved.
userFCs	Data from a user performed screen: A data frame with the same format of the R ojbects composing the reference dataset, i.e. first two columns containing sgRNAs' identifiers and HGNC symbols of targete gene (headers = sgRNA and gene, respectively), followed by one column per screen replicate containing sgRNAs' fold-changes.
stats	Boolean parameter specifying whether to show average statistics. Default is TRUE.
saveToFig	Boolean parameter specifying whether to save all figures in PDF format. Default is TRUE.
display	Boolean parameter specifying whether to display graphics on R console. Default is FALSE.

Details

The function computes sgRNAs logFCs' statistics such as average range, median, interquartile range, 10th-90th percentile range, skewness and kurtosis across the six HT-29 screens plus the User-defined screen, if provided.

Author(s)

Raffaele Iannuzzi, Ichcha Manipur, Francesco Iorio

References

[1] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

See Also

H29R.downloadRefData

Examples

```
## Creating a temporary folder to store the HT29 reference dataset
## and the pdf created by this function
dir.create('tmpDir')

## Downloading the HT29 reference dataset in the temporary folder
HT29R.downloadRefData(destFolder = 'tmpdir')

## Loading CRISPRcleanR library to use example screen data
library(CRISPRcleanR)

## Deriving the path of the file with the example dataset,
## from the mutagenesis of the HT-29 colorectal cancer cell line
fn <- dir('tmpDir')
fn <- grep('_foldChanges.Rdata',fn,value=TRUE)
```

```
## Loading library Annotation
data('KY_Library_v1.0')

## Loading, median-normalizing and computing fold-changes for the example dataset
normANDfcs<-ccr.NormfoldChanges(fn,min_reads=30,EXPname='ExampleScreen',
                                libraryAnnotation = KY_Library_v1.0,
                                display = FALSE)

ExampleScreen<-normANDfcs$logFCs

## Evaluating screen reproducibility of HT29 reference and user defined data
## both, using Project Score criteria
HT29R.FCdistributions(refDataDir='tmpDir',
                      resDir='tmpDir',
                      userFCs=ExampleScreen,
                      stats=TRUE,
                      saveToFig=TRUE,
                      display=FALSE)

## Checking results
system2('open', args = 'tmpDir/QC_FCdistproperties.pdf', wait = FALSE)


## Removing Example dataset processed files
file.remove('ExampleScreen_foldChanges.Rdata')
file.remove('ExampleScreen_normCounts.Rdata')
```

`HT29R.prSCORE_rCorr` *Pair-wise screen replicate correlations and background correlations from Project Score.*

Description

Correlation scores obtained by comparing sgRNA essentiality profiles, considering KY-Library[1] specific informative/reproducible sgRNAs depletion fold-changes between replicates of the same experiment, and between all possible pairs of individual replicates across experiments, from Project Score[2]].

Usage

```
data("HT29R.prSCORE_rCorr")
```

Format

A list of two numerical vectors:

`BGscores` a numeric vector with 441387 entries, containing the correlation scores obtained by comparing profiles of KY-Library[1] specific informative/reproducible sgRNAs depletion fold-changes between all possible pairs of individual replicates across experiments, i.e. background correlation.

`REPscores` a numeric vector with 883 entries, containing the correlation scores obtained by comparing profiles of KY-Library[1] specific informative/reproducible sgRNAs depletion fold-changes between replicates of the same experiment.

References

[1] Tzelepis K, Koike-Yusa H, De Braekeleer E, Li Y, Metzakopian E, Dovey OM, et al. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. Cell Rep. 2016;17:1193–205.

[2] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

Examples

```
library(CRISPRcleanR)
data(HT29R.prSCORE_rCorr)

ccr.multDensPlot(
  list(density(HT29R.prSCORE_rCorr$BGscores),
       density(HT29R.prSCORE_rCorr$REPscores)),
  XLIMS = c(0,1),
 TITLE = 'Project Score Observed vs. Expected replicate correlations\n (informative sgRNAs only)',
  COLS = c('gray','darkgreen'),LEGentries = c('expected','observed'),XLAB='R')
```

HT29R.GL_prSCORE_rCorr

Gene level pair-wise screen replicate correlations and background correlations from Project Score.

Description

Correlation scores obtained by comparing gene essentiality profiles, considering KY-Library[1] specific informative/reproducible sgRNAs depletion fold-changes (averaged on a targeted gene basis) between replicates of the same experiment, and between all possible pairs of individual replicates across experiments, from Project Score[2]].

Usage

```
data("HT29R.GL_prSCORE_rCorr")
```

Format

A list of two numerical vectors:

`BGscores` a numeric vector with 441387 entries, containing the correlation scores obtained by comparing profiles of KY-Library[1] specific informative/reproducible sgRNAs depletion fold-changes (averaged on a targeted gene basis) between all possible pairs of individual replicates across experiments, i.e. background correlation.

`REPscores` a numeric vector with 883 entries, containing the correlation scores obtained by comparing profiles of KY-Library[1] specific informative/reproducible sgRNAs depletion fold-changes (avereged on a targeted gene basis) between replicates of the same experiment.

References

[1] Tzelepis K, Koike-Yusa H, De Braekeleer E, Li Y, Metzakopian E, Dovey OM, et al. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. Cell Rep. 2016;17:1193–205.

[2] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

Examples

```
library(CRISPRcleanR)
data(HT29R.GL_prSCORE_rCorr)

ccr.multDensPlot(
  list(density(HT29R.GL_prSCORE_rCorr$BGscores),
       density(HT29R.GL_prSCORE_rCorr$REPscores)),
  XLIMS = c(0,1),
 TITLE = 'Project Score Observed vs. Expected replicate correlations\n (Gene level depletion fc, informative sg
  COLS = c('gray','darkgreen'),LEGentries = c('expected','observed'),XLAB='R')
```

HT29R.evaluateReps *Low-level quality control (QC) analysis: reproducibility across technical replicates.*

Description

Assessing CRISPR-Cas9 genome-wide pooled screens' reproducibility using Project Score metrics and criteria.

Usage

```
HT29R.evaluateReps(refDataDir='./',
                   resDir='./',
                   userFCs=NULL,
                   geneLevel=TRUE,
                   saveToFig=TRUE,
                   display=FALSE)
```

Arguments

refDataDir	Reference HT29 dataset directory: a string specifying the location of the processed HT29 reference dataset.
resDir	Output directory: a string specifying the directory where the output of this function (a pdf file with multiple plots) should be saved.
userFCs	Data from a user performed screen: A data frame with the same format of the R ojbects composing the reference dataset, i.e. first two columns containing sgRNAs' identifiers and HGNC symbols of targete gene (headers = sgRNA and gene, respectivley), followed by one column per screen replicate containing sgRNAs' fold-changes.
geneLevel	Boolean parameter specifying whether the sgRNA profiles of depletion fold-changes should be collapsed by averaging them on a targeted gene basis. Default value = TRUE.

saveToFig	Boolean parameter specifying whether to save all figures in PDF format. Default value = TRUE.
display	Boolean parameter specifying whether to display graphics on R console. Default value = FALSE.

Details

This function computes correlation scores between each pair of screen replicates for the HT29 reference dataset as well as for used defined data. This is performed as for the Project Score data [1]: considering only a set of 838 most informative sgRNAs (in the `HT29R.reproducible_GeneGuides object`), defined as those targeting the same genes and with an average pairwise Pearson's correlation > 0.6 between corresponding patterns of depletion fold-changes (FCs) across hundreds of screened cell lines. Per construction, the depletion patterns of these sgRNAs are both reproducible and informative (as they involve genes carrying an actual fitness signal). Computing correlation scores between replicates of the same screen on the domain of these sgRNAs only allows estimating a null distribution of replicate correlations and computing a reproducibility threshold defined as the minimal correlation score that should be observed between replicates of the same screen (R = 0.68, using the `HT29R.GL_prSCORE_rCorr_Reprod` object). This is performed as genome-wide correlation scores computed between replicates of the same CRISPR-Cas9 pooled genome-wide viability screen are generally always very high and indistiguishable from expectation due to only a small percentage of genes exerting an effect on cellular fitness upon knock-out.

This function produces a pdf file showing the plot of the expected/observed distributions of replicate pair-wise correlation scores and reproducibility threshold from the Project Score dataset. Below this plots, the pair-wise correlation scores computed between replicates of the HT29 reference dataset are also plotted. If the user provides its own data (trough the `userFCs` parameter) then pair-wise correlation scores computed between replicates of this screens are overlaied on the kernel distributions. In this case, a specification of the number of replicates reaching threshold is printed in the console.

Author(s)

Ichcha Manipur, Raffaele Iannuzzi, Francesco Iorio (fi1@sanger.ac.uk)

References

[1] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

See Also

`H29R.downloadRefData`, `HT29R.reproducible_GeneGuides`, `HT29R.prSCORE_rCorr`, `HT29R.GL_prSCORE_rCorr`

Examples

```
## Creating a temporary folder to store the HT29 reference dataset
## and the pdf created by this function
dir.create('tmpDir')

## Downloading the HT29 reference dataset in the temporary folder
HT29R.downloadRefData(destFolder = 'tmpDir')

## Loading CRISPRcleanR library to use example screen data
library(CRISPRcleanR)
```

```
## Deriving the path of the file with the example dataset,
## from the mutagenesis of the HT-29 colorectal cancer cell line
fn<-paste(system.file('extdata', package = 'CRISPRcleanR'),'/HT-29_counts.tsv',sep='')

## Loading library Annotation
data('KY_Library_v1.0')

## Loading, median-normalizing and computing fold-changes for the example dataset
normANDfcs<-ccr.NormfoldChanges(fn,min_reads=30,EXPname='ExampleScreen',
                                libraryAnnotation = KY_Library_v1.0,
                                display = FALSE)

ExampleScreen<-normANDfcs$logFCs

## Evaluating screen reproducibility of HT29 reference and user defined data
## both, using Project Score criteria
HT29R.evaluateReps(refDataDir='tmpDir',
                   resDir='tmpDir',
                   userFCs=ExampleScreen,
                   geneLevel=TRUE,
                   saveToFig=TRUE,
                   display=FALSE)

## Checking results
system2('open', args = 'tmpDir/QC_REPRODUCIBILITY_GENElevel.pdf',sep=""), wait = FALSE)

## Removing Example dataset processed files
file.remove('ExampleScreen_foldChanges.Rdata')
file.remove('ExampleScreen_normCounts.Rdata')
```

`HT29R.consensus_GeneGuides` *Library specific sgRNAs' consensus across HT29 screens.*

Description

86875 KY-library[1] specific sgRNAs for evaluating CRISPR-Cas9 pooled genome-wide viability screen replicates.

Usage

```
data(HT29R.consensus_GeneGuides)
```

Format

A vector of strings with entries corresponding to sgRNAs' identifiers.

Details

This file contain a list of sgRNAs strictly shared by each HT29 genome-wide CRISPR-Cas9 depletion screen and it's used by the `HT29R.expSimilarity` function to evaluate the User-provided screen (i.e., the intersection between consensus and User's sgRNAs is considered for comparing screens).

References

[1] Tzelepis K, Koike-Yusa H, De Braekeleer E, Li Y, Metzakopian E, Dovey OM, et al. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. Cell Rep. 2016;17:1193–205.

[2] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

Examples

```
data(HT29R.consensus_GeneGuides)
head(HT29R.consensus_GeneGuides)
```

HT29R.reproducible_GeneGuides

Library specific informative/reproducible sgRNAs.

Description

838 KY-library[1] specific informative/reproducible sgRNAs (targeting 308 genes) for evaluating CRISPR-Cas9 pooled genome-wide viability screen replicates.

Usage

```
data(HT29R.reproducible_GeneGuides)
```

Format

A vector of strings with entries corresponding to sgRNAs' identifiers.

Details

Genome-wide correlation scores computed between replicates of the same CRISPR-Cas9 pooled genome-wide viability screen are generally always very high and indistiguishable from expectation due to only a small percentage of genes exerting an effect on cellular fitness upon knock-out. In [2] we have selected a set of 838 most informative sgRNAs, defined as those targeting the same genes and with an average pairwise Pearson's correlation > 0.6 between corresponding patterns of depletion fold-changes (FCs) across hundreds of screened cell lines. Per construction, the depletion patterns of these sgRNAs are both reproducible and informative (as they involve genes carrying an actual fitness signal). Computing correlation scores between replicates of the same screen on the domain of these sgRNAs only allowed the estimation of a null distribution of replicate correlations and computing a reproducibility threshold defined as the minimal correlation score that should be observed between replicates of the same screen (R = 0.68).

References

[1] Tzelepis K, Koike-Yusa H, De Braekeleer E, Li Y, Metzakopian E, Dovey OM, et al. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. Cell Rep. 2016;17:1193–205.

[2] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

Examples

```
data(HT29R.reproducible_GeneGuides)
head(HT29R.reproducible_GeneGuides)
```

HT29R.prSCORE_bkgr_screen_similarity_HI

Pair-wise screen replicate correlations (i.e., background correlations) from Project Score considering most-informative sgRNAs.

Description

Correlation scores obtained by comparing genes' essentiality profiles (i.e., genes' depletion log fold-changes) between all possible pairs of sample (325 x 325), considering the Project Score[1] data matrix for 308 genes x 325 samples (= 324 cell lines + HT29v1.1).

Usage

```
data("HT29R.prSCORE_bkgr_screen_similarity_HI")
```

Format

A numerical vector of Pearson's correlation coefficents.

Details

A numerical vector with 52650 entries, containing the correlation scores obtained by comparing profiles of 308 genes' depletion fold-changes between all possible pairs of individual experiment (i.e., the "HI" gene-level background correlation). See function `HT29R.expSimilarity` and `HT29R.reproducible_GeneGuides` descriptions in reference the manual for further details.

References

[1] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

Examples

```
library(HT29benchmark)
data(HT29R.prSCORE_bkgr_screen_similarity_HI)

plot(density(HT29R.prSCORE_bkgr_screen_similarity_HI),
  xlim=c(0,1),
  main="Project Score background distribution\n (308 most-informative genes)",
  col="grey",
  lwd=2,
  xlab="R")
```

HT29R.prSCORE_bkgr_screen_similarity_sgRNA_HI

Pair-wise screen replicate correlations (i.e., background correlations) from Project Score considering most-informative sgRNAs.

Description

Correlation scores obtained by comparing genes' essentiality profiles (i.e., genes' depletion log fold-changes) between all possible pairs of sample (325 x 325), considering the Project Score[1] data matrix for 838 KY-library[1] specific informative/reproducible sgRNAs x 325 samples (= 324 cell lines + HT29v1.1).

Usage

```
data("HT29R.prSCORE_bkgr_screen_similarity_sgRNA_HI")
```

Format

A numerical vector of Pearson's correlation coefficents.

Details

A numerical vector with 52650 entries, containing the correlation scores obtained by comparing profiles of 838 sgRNAs' depletion fold-changes between all possible pairs of individual experiment (i.e., the "HI" guide-level background correlation). See function `HT29R.expSimilarity` and `HT29R.reproducible_GeneGuides` descriptions in the reference manual for further details.

References

[1] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

Examples

```
library(HT29benchmark)
data(HT29R.prSCORE_bkgr_screen_similarity_sgRNA_HI)

plot(density(HT29R.prSCORE_bkgr_screen_similarity_sgRNA_HI),
  xlim=c(-0.2,1),
  main="Project Score background distribution\n (838 most-informative sgRNAs)",
  col="grey",
  lwd=2,
  xlab="R")
```

HT29R.prSCORE_bkgr_screen_similarity_sgRNA *Pair-wise screen replicate correlations (i.e., background correlations) from Project Score.*

Description

Correlation scores obtained by comparing genes' essentiality profiles (i.e., sgRNAs' depletion log fold-changes) between all possible pairs of sample (325 x 325), considering the Project Score[1] data matrix for 86875 KY-library[1] specific sgRNAs x 325 samples (= 324 cell lines + HT29v1.1).

Usage

```
data("HT29R.prSCORE_bkgr_screen_similarity_sgRNA")
```

Format

A numerical vector of Pearson's correlation coefficents.

Details

A numerical vector with 52650 entries, containing the correlation scores obtained by comparing profiles of 86875 sgRNAs' depletion fold-changes between all possible pairs of individual experiment (i.e., the "All" guide-level background correlation). See function `HT29R.expSimilarity` description in the reference manual for further details.

References

[1] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

Examples

```
library(HT29benchmark)
data(HT29R.prSCORE_bkgr_screen_similarity_sgRNA)

plot(density(HT29R.prSCORE_bkgr_screen_similarity_sgRNA),
  xlim=c(0,1),
  main="Project Score background distribution\n (86875 sgRNAs)",
  col="grey",
  lwd=2,
  xlab="R")
```

HT29R.prSCORE_bkgr_screen_similarity

Pair-wise screen replicate correlations (i.e., background correlations) from Project Score.

Description

Correlation scores obtained by comparing genes' essentiality profiles (i.e., genes' depletion log fold-changes) between all possible pairs of sample (325 x 325), considering the Project Score[1] data matrix for 17995 genes x 325 samples (= 324 cell lines + HT29v1.1).

Usage

```
data("HT29R.prSCORE_bkgr_screen_similarity")
```

Format

A numerical vector of Pearson's correlation coefficents.

Details

A numerical vector with 52650 entries, containing the correlation scores obtained by comparing profiles of 17995 genes' depletion fold-changes between all possible pairs of individual experiment (i.e., the "All" gene-level background correlation). See function `HT29R.expSimilarity` description in the reference manual for further details.

References

[1] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

Examples

```
library(HT29benchmark)
data(HT29R.prSCORE_bkgr_screen_similarity)

plot(density(HT29R.prSCORE_bkgr_screen_similarity),
  xlim=c(0,1),
  main="Project Score background distribution\n (17995 targeted genes)",
  col="grey",
  lwd=2,
  xlab="R")
```

HT29R.expSimilarity *Low-level quality control (QC) analysis: quantifying screens' similarity.*

Description

Assessing CRISPR-Cas9 genome-wide pooled screens' comparability using Project Score data.

Usage

```
HT29R.expSimilarity(refDataDir='./',
                    resDir='./',
                    userFCs=NULL,
                    geneGuides=c("All","HI"),
                    geneLevel=TRUE,
                    Rscores=TRUE,
                    saveToFig=TRUE,
                    display=FALSE)
```

Arguments

refDataDir	Reference HT29 dataset directory: a string specifying the location of the processed HT29 reference dataset.
resDir	Output directory: a string specifying the directory where the output of this function (a pdf file with multiple plots) should be saved.
userFCs	Data from a user performed screen: A data frame with the same format of the R ojbects composing the reference dataset, i.e. first two columns containing sgRNAs' identifiers and HGNC symbols of targete gene (headers = sgRNA and gene, respectivley), followed by one column per screen replicate containing sgRNAs' fold-changes.
geneGuides	String parameter specifying whether to use all sgRNAs (genome-wide) or just the most reproducible and informative sgRNAs (n=838)
geneLevel	Boolean parameter specifying whether the sgRNA profiles of depletion fold-changes should be collapsed by averaging them on a targeted gene basis. Default value = TRUE.
Rscores	Boolean parameter specifying whether to compute correlation scores matrix between averaged screen replicates for the HT29 reference dataset as well as for user defined data. Default value = TRUE.
saveToFig	Boolean parameter specifying whether to save all figures in PDF format. Default value = TRUE.
display	Boolean parameter specifying whether to display graphics on R console. Default value = FALSE.

Details

This function computes correlation scores between each averaged screens' replicate across the HT29 reference dataset and between each averaged HT29 screens' replicate and user defined data (through the userFCs parameter) in order to measure depletion fold-changes (FCs) similarities across the HT29 screens and the user defined depletion fold-changes patterns. The analysis can be

performed considering the set of 838 most informative sgRNAs only (named HT29R.reproducible_GeneGuides as seen for the HT29R.evaluateReps function, or considering the sgRNAs consensus (named HT29R.consensus_GeneGuides) across the reference and for the user defined screen. If the geneLevel parameter is set to TRUE, HGNC genes' symbols will be used instead of sgRNAs identifiers. The background distributions will be chosen based on the combination of parameters selected (e.g., if geneGuides parameter is set to "All" and geneLevel is set to TRUE, HT29R.prSCORE_bkgr_screen_similarity will be selected instead of {HT29R.prSCORE_bkgr_screen_similarity_sgRNA

Author(s)

Raffaele Iannuzzi, Ichcha Manipur, Francesco Iorio (fi1@sanger.ac.uk)

References

[1] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

See Also

HT29R.downloadRefData, HT29R.reproducible_GeneGuides, HT29R.consensus_GeneGuides, HT29R.prSCORE_bkgr_screen_similarity, HT29R.prSCORE_bkgr_screen_similarity_HI, HT29R.prSCORE_bkgr_
HT29R.prSCORE_bkgr_screen_similarity_sgRNA_HI

Examples

```
## Creating a temporary folder to store the HT29 reference dataset
## and the pdf created by this function
dir.create('tmpDir')

## Downloading the HT29 reference dataset in the temporary folder
HT29R.downloadRefData(destFolder = 'tmpDir')

## Loading CRISPRcleanR library to use example screen data
library(CRISPRcleanR)

## Loading library Annotation
data('KY_Library_v1.0')

## Deriving the path of the file with the example dataset,
## from the mutagenesis of the HT-29 colorectal cancer cell line
fn<-paste(system.file('extdata', package = 'CRISPRcleanR'),'/HT-29_counts.tsv',sep='')

## Loading, median-normalizing and computing fold-changes for the example dataset
normANDfcs<-ccr.NormfoldChanges(fn,min_reads=30,EXPname='ExampleScreen',
                                libraryAnnotation = KY_Library_v1.0,
                                display = FALSE)

ExampleScreen<-normANDfcs$logFCs

## Evaluating screen similarity of HT29 reference and user defined data
## both, using Project Score criteria

HT29R.expSimilarity(refDataDir='tmpDir',
                    resDir='tmpDir',
                    userFCs=ExampleScreen,
                    geneGuides='All',
```

```
                    geneLevel=TRUE,
                    Rscore=TRUE,
                    saveToFig=TRUE,
                    display=FALSE)

## Checking results
system2('open', args = 'tmpDir/All_QC_SIMILARITY_GENElevel.pdf', wait = FALSE)
system2('open', args = 'tmpDir/All_SCATTERPLOT_R_MATRIX_GENElevel.pdf', wait = FALSE)

## Removing Example dataset processed files
file.remove('ExampleScreen_foldChanges.Rdata')
file.remove('ExampleScreen_normCounts.Rdata')
```

`HT29R.PhenoIntensity` *High-level quality control (QC) analysis using Glass's Delta scores.*

Description

Assessing screens' performance in recovering gene essentiality profiles using sets of known essential (True Positives, TP) and known non-essential (True Negatives, TN) genes.

Usage

```
HT29R.PhenoIntensity(refDataDir='./',
                     resDir='./',
                     userFCs=NULL,
                     geneLevel=TRUE,
                     saveToFig=TRUE,
                     display=FALSE)
```

Arguments

`refDataDir`	Reference HT29 dataset directory: a string specifying the location of the processed HT29 reference dataset.
`resDir`	Output directory: a string specifying the directory where the output of this function (a pdf file with multiple plots) should be saved.
`userFCs`	Data from a user performed screen: A data frame with the same format of the R ojbects composing the reference dataset, i.e. first two columns containing sgRNAs' identifiers and HGNC symbols of targete gene (headers = sgRNA and gene, respectivley), followed by one column per screen replicate containing sgRNAs' fold-changes.
`geneLevel`	Boolean parameter specifying whether the sgRNA profiles of depletion fold-changes should be collapsed by averaging them on a targeted gene basis. Default value = TRUE.
`saveToFig`	Boolean parameter specifying whether to save all figures in PDF format. Default value = TRUE.
`display`	Boolean parameter specifying whether to display graphics on R console. Default value = FALSE.

Details

This function computes Glass's Delta scores as a metric to quantify screen quality by leveraging the intensity of the phenotype exerted by the inactivation of prior-known essential (E), ribosomal (R) and non-essential (N) genes. These scores account for the difference between the average depletion FCs of the genes in E (respectively R) and that of genes known to be non-essential (i.e., genes that are rarely expressed across cell lines 3) (N) in relation to the standard deviation of the depletion FCs of the genes in E (respectively R), as it follows: Glass's Delta(X) score = | mu(FC(x belongs to X)) - mu(FC(n belongs to N)) | / sigma(FC(x belongs to X)) where X belongs to E, R , mu, sigma indicate mean and standard deviation, respectively. This function produces a pdf file (named `'ALLSCREENS_PHENOINTENSITY.pdf'` in the specified directory with multiple plots for each HT29 screen of the observed distributions of prior-known essential (E), ribosomal (R) and non-essential (N) genes coupled with the respective Glass's Delta values. If provided through the `userFCs` parameter, the function produces a second pdf file (named `'REFERENCE_PHENOINTENSITY.pdf'`) showing the same distributions for the user defined data. Boxplots summarizing the Glass's Delta score distributions of the two set of essential genes (i.e., Ribosomal protein genes (R) and Other essential genes (E)) across the six HT29 screens plus for the user defined screen (if provided) are also produced.

Author(s)

Ichcha Manipur & Francesco Iorio (fi1@sanger.ac.uk)

References

[1] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

See Also

`EssGenes.ribosomalProteins`, `BAGEL_essential`, `BAGEL_nonEssential`

Examples

```
## Creating a temporary folder to store the HT29 reference dataset
## and the pdf created by this function
dir.create('tmpdir')

## Downloading the HT29 reference dataset in the temporary folder
HT29R.downloadRefData(destFolder = 'tmpdir')

## Loading CRISPRcleanR library to use example screen data
library(CRISPRcleanR)

## Deriving the path of the file with the example dataset,
## from the mutagenesis of the HT-29 colorectal cancer cell line
fn<-paste(system.file('extdata', package = 'CRISPRcleanR'),'/HT-29_counts.tsv',sep='')

## Loading library Annotation
data('KY_Library_v1.0')

## Loading, median-normalizing and computing fold-changes for the example dataset
normANDfcs<-ccr.NormfoldChanges(fn,min_reads=30,EXPname='ExampleScreen',
                                libraryAnnotation = KY_Library_v1.0,
                                display = FALSE)
```

```
ExampleScreen<-normANDfcs$logFCs

## Evaluating phenotype intensity of HT29 reference and user defined data
HT29R.PhenoIntensity(refDataDir = 'tmpDir',
                     resDir = 'tmpDir',
                     userFCs = ExampleScreen,
                     geneLevel=TRUE,
                     saveToFig = TRUE,
                     display=FALSE)

## Checking results
system2('open', args = 'tmpDir/ALLSCREENS_PHENOINTENSITY.pdf', wait = FALSE)
system2('open', args = 'tmpDir/REFERENCE_PHENOINTENSITY.pdf', wait = FALSE)

## Removing Example dataset processed files
file.remove('ExampleScreen_foldChanges.Rdata')
file.remove('ExampleScreen_normCounts.Rdata')
```

`HT29R.ROCanalysis` *High-level quality control (QC) analysis using AUROC and AUPRRC metrics.*

Description

Assessing screens' performance in recovering gene essentiality profiles using sets of known essential (True Positives, TP) and known non-essential (True Negatives, TN) genes based on their sgRNA logFC rank positions while computing and visualising ROC and PrRC curves .

Usage

```
HT29R.ROCanalysis(refDataDir='./',
                  resDir='./',
                  positives,
                  negatives,
                  userFCs=NULL,
                  geneLevel=TRUE,
                  saveToFig=TRUE,
                  display=FALSE)
```

Arguments

`refDataDir`	Reference HT29 dataset directory: a string specifying the location of the processed HT29 reference dataset.
`resDir`	Output directory: a string specifying the directory where the output of this function (a pdf file with multiple plots) should be saved.
`positives`	A vector of strings containing a reference set of positive cases: HGNC symbols of essential genes or identifiers of their targeting sgRNAs. This must be a subset disjointed from negatives.
`negatives`	A vector of strings containing a reference set of negative cases: HGNC symbols of essential genes or identifiers of their targeting sgRNAs. This must be a subset disjointed from positives.

userFCs	Data from a user performed screen: A data frame with the same format of the R ojbects composing the reference dataset, i.e. first two columns containing sgRNAs' identifiers and HGNC symbols of targete gene (headers = sgRNA and gene, respectivley), followed by one column per screen replicate containing sgRNAs' fold-changes.
geneLevel	Boolean parameter specifying whether the sgRNA profiles of depletion fold-changes should be collapsed by averaging them on a targeted gene basis. Default value = TRUE.
saveToFig	Boolean parameter specifying whether to save all figures in PDF format. Default value = TRUE.
display	Boolean parameter specifying whether to display graphics on R console. Default value = FALSE.

Details

This function computes AUROC and AUPRRC in order to assess the ability of each HT29 screen plus for the user defined screen (if provided through the `userFCs` parameter) to detect prior-known essential (positives) and non-essential (negatives) set of genes using ranked depletion fold-changes (FCs) values as predictors. This function produces a pdf file (named `'ALLSCREENS_ROCs.pdf'` in the specified directory with multiple plots showing ROC and PrRc curves for each HT29 screen. If provided through the `userFCs` parameter, the function produces a second pdf file (named `'USER_ROCs.pdf'`) with ROC and PrRc curves of the user defined data. If `display` is set to TRUE, a plot showing the overlayed curves' profiles will be displayed on R console.

Author(s)

Ichcha Manipur & Francesco Iorio (fi1@sanger.ac.uk)

References

[1] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

See Also

`HT29R.PhenoIntensity`, `HT29R.FDRconsensus`, `ccr.ROC_Curve`, `ccr.PrRc_Curve`, `BAGEL_essential`, `BAGEL_nonEssential`

Examples

```
## Creating a temporary folder to store the HT29 reference dataset
## and the pdf created by this function
dir.create('tmpDir')

## Downloading the HT29 reference dataset in the temporary folder
HT29R.downloadRefData(destFolder = 'tmpDir')

## Loading CRISPRcleanR library to use example screen data
library(CRISPRcleanR)

## Deriving the path of the file with the example dataset,
## from the mutagenesis of the HT-29 colorectal cancer cell line
fn<-paste(system.file('extdata', package = 'CRISPRcleanR'),'/HT-29_counts.tsv',sep='')
```

```
## Loading library Annotation
data('KY_Library_v1.0')

## Loading, median-normalizing and computing fold-changes for the example dataset
normANDfcs<-ccr.NormfoldChanges(fn,min_reads=30,EXPname='ExampleScreen',
                                libraryAnnotation = KY_Library_v1.0,
                                display = FALSE)

ExampleScreen<-normANDfcs$logFCs

## deriving sgRNAs targeting essential and non-essential genes (respectively)
BAGEL_essential_sgRNAs<-ccr.genes2sgRNAs(KY_Library_v1.0,BAGEL_essential)
BAGEL_nonEssential_sgRNAs<-ccr.genes2sgRNAs(KY_Library_v1.0,BAGEL_nonEssential)

## Computing classification performances at the sgRNA level
HT29R.ROCanalysis(refDataDir = 'tmpDir',
                      resDir = 'tmpDir',
                      positives = BAGEL_essential_sgRNAs
                      negatives = BAGEL_nonEssential_sgRNAs
                      userFCs = ExampleScreen,
                      geneLevel=FALSE,
                      saveToFig = TRUE,
                      display=FALSE)

## Checking results
system2('open', args = 'tmpDir/ALLSCREENS_ROCs.pdf', wait = FALSE)
system2('open', args = 'tmpDir/USER_ROCs.pdf', wait = FALSE)

## Removing Example dataset processed files
file.remove('ExampleScreen_foldChanges.Rdata')
file.remove('ExampleScreen_normCounts.Rdata')
```

`HT29R.FDRconsensus` *Defining a positive consensus of essential gene across the HT29 screens using the FDR ranking method.*

Description

The FDR method is applied for each HT29 screen, then a strict intersection of genes classified as "essential" at a defined FDR threshold (0.05 by default) is taken. The positive consensus of genes identified is used as a probe to measure the distance (Cohen's or Glass's Delta) between the depletion fold-changes (FCs) of the positive and the negative consensus of genes for each screen, which is also computed by the same function. The function uses a set of prior-known essential and non essential genes (respectively the `BAGEL_essential` and `BAGEL_nonEssential`) as reference.

Usage

```
HT29R.FDRconsensus(refDataDir="./",
                   resDir="./",
                   userFCs=NULL,
                   distance=c("GlDelta","Cohen's"),
                   FDRth=0.05,
                   saveToFig=TRUE,
                   display=FALSE)
```

Arguments

refDataDir	Reference HT29 dataset directory: a string specifying the location of the processed HT29 reference dataset.
resDir	Output directory: a string specifying the directory where the output of this function (a pdf file with multiple plots) should be saved.
userFCs	Data from a user performed screen: A data frame with the same format of the R ojbects composing the reference dataset, i.e. first two columns containing sgRNAs' identifiers and HGNC symbols of targeted gene (headers = sgRNA and gene, respectivley), followed by one column per screen replicate containing sgRNAs' fold-changes.
distance	Type of distance the user want to measure between the two consensus: A character string indicating "Cohen's" for the Cohen's d or "GlDelta" for Glass's Delta.
FDRth	If different from NULL, will be a numerical value >=0 and <=1 specifying the false discovery rate threshold at which fixed recall will be computed. Default value = 0.05.
saveToFig	Boolean parameter specifying whether to save all figures in PDF format. Default value = TRUE.
display	Boolean parameter specifying whether to display graphics on R console. Default value = FALSE.

Details

This function produces a pdf file (named 'FDR_CONSENSUS_DIST.pdf' in the specified directory with multiple plots showing the depletion fold-changes (FCs) of the positive and negative consensus of genes for each HT29 screen plus for the user defined screen if provided through the userFCs parameter. A second plot shows the distribution of the distances computed for each screen. If data are provided by the user, it allows to measure how far the HT29-specific genes log fold-changes are from the reference.

Author(s)

Raffaele Iannuzzi & Francesco Iorio (fi1@sanger.ac.uk)

References

[1] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

See Also

HT29R.PhenoIntensity, HT29R.ROCanalysis, BAGEL_essential, BAGEL_nonEssential

Examples

```
## Creating a temporary folder to store the HT29 reference dataset
## and the pdf created by this function
dir.create('tmpDir')

## Downloading the HT29 reference dataset in the temporary folder
HT29R.downloadRefData(destFolder = 'tmpDir')

## Loading CRISPRcleanR library to use example screen data
```

```
library(CRISPRcleanR)

## Deriving the path of the file with the example dataset,
## from the mutagenesis of the HT-29 colorectal cancer cell line
fn<-paste(system.file('extdata', package = 'CRISPRcleanR'),'/HT-29_counts.tsv',sep='')

## Loading library Annotation
data('KY_Library_v1.0')

## Loading, median-normalizing and computing fold-changes for the example dataset
normANDfcs<-ccr.NormfoldChanges(fn,min_reads=30,EXPname='ExampleScreen',
                                libraryAnnotation = KY_Library_v1.0,
                                display = FALSE)

ExampleScreen<-normANDfcs$logFCs

## Compute 5 percent FDR consensus
RES <- HT29R.FDRconsensus(refDataDir = 'tmpDir',
                      resDir = 'tmpDir',
                      userFCs = ExampleScreen,
                      distance = "Cohen's",
                      FDRth = 0.05,
                      saveToFig = TRUE,
                      display=FALSE)

## Checking results
system2('open', args = 'tmpDir/FDR_CONSENSUS_DIST.pdf', wait = FALSE)

## Removing Example dataset processed files
file.remove('ExampleScreen_foldChanges.Rdata')
file.remove('ExampleScreen_normCounts.Rdata')
```

`HT29R.runCrisprQC_Analysis` *Wrapper function: download reference data and run all QC analysis from low-level to high-level.*

Usage

```
HT29R.runCrisprQC_Analysis(data = c("FCs","rawCounts") ,
                           userFCs=NULL,
                           positives,
                           negatives,
                           outdir="./")
```

Arguments

`data`	String parameter specifying what type of data to dowload. Possible values are "rawCounts" for plain .tsv files containing raw sgRNA countsare or "FCs" (default) for R objects containing sgRNA normalised depletion fold-changes: data frames in which the first two columns contain sgRNAs' identifiers and HGNC symbols of targeted gene, followed by one column per screen replicate containing sgRNAs' fold-changes.

userFCs	Data from a user performed screen: A data frame with the same format of the R ojbects composing the reference dataset, i.e. first two columns containing sgRNAs' identifiers and HGNC symbols of targete gene (headers = sgRNA and gene, respectivley), followed by one column per screen replicate containing sgRNAs' fold-changes.
positives	A vector of strings containing a reference set of positive cases: HGNC symbols of essential genes or identifiers of their targeting sgRNAs. This must be a subset disjointed from negatives.
negatives	A vector of strings containing a reference set of negative cases: HGNC symbols of essential genes or identifiers of their targeting sgRNAs. This must be a subset disjointed from positives.
outdir	Output directory: a string specifying the directory where the output of this function (a pdf file with multiple plots) should be saved

Details

The function create a folder named "PLOTS" in which the figures will be stored. The code will ask the user to provide two parameters whose are necessary for the low-level QC analysis. The first define the subset of guides the user intends to use (i.e., "All" for the sgRNAs targeting all genes in the genome-wide library and "HI" for the sgRNAs targeting only the most-informative and highly-reproducible ones as described in [1]), the second define the level of inspection (i.e., TRUE if inspecting at the gene-level or FALSE if at the guide-level).

Author(s)

Raffaele M. Iannuzzi & Francesco Iorio (fi1@sanger.ac.uk)

References

[1] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511–6.

See Also

HT29R.downloadRefData, HT29R.evaluateReps, HT29R.expSimilarity, HT29R.PhenoIntensity, HT29R.ROCanalysis, HT29R.FDRconsensus, BAGEL_essential, BAGEL_nonEssential

Examples

```
## Creating a temporary folder to store the HT29 reference dataset
## and the pdf created by this function
dir.create('tmpDir')

## Downloading the HT29 reference dataset in the temporary folder
HT29R.downloadRefData(destFolder = 'tmpDir')

## Loading CRISPRcleanR library to use example screen data
library(CRISPRcleanR)

## Deriving the path of the file with the example dataset,
## from the mutagenesis of the HT-29 colorectal cancer cell line
fn<-paste(system.file('extdata', package = 'CRISPRcleanR'),'/HT-29_counts.tsv',sep='')

## Loading library Annotation
```

```
data('KY_Library_v1.0')

## Loading, median-normalizing and computing fold-changes for the example dataset
normANDfcs<-ccr.NormfoldChanges(fn,min_reads=30,EXPname='ExampleScreen',
                              libraryAnnotation = KY_Library_v1.0,
                              display = FALSE)

ExampleScreen<-normANDfcs$logFCs

## Run all with interactive wrapper

HT29R.runCrisprQC_Analysis(data = "FCs",
                           userFCs = ExampleScreen,
                           positives = BAGEL_essential,
                           negatives = BAGEL_nonEssential,
                           outdir = 'tmpDir')

## Checking results
system2('open', args = 'tmpDir/PLOTS/', wait = FALSE)

## Removing Example dataset processed files
file.remove('ExampleScreen_foldChanges.Rdata')
file.remove('ExampleScreen_normCounts.Rdata')
```