Package 'HPAStainR'

March 30, 2021

Title Queries the Human Protein Atlas Staining Data for Multiple
Proteins and Genes

Version 1.0.3

Description This package is built around the HPAStainR function. The purpose of
the HPAStainR function is to query the visual staining data in the Human
Protein Atlas to return a table of staining ranked cell types. The function
also has multiple arguements to personalize to output as well to include
cancer data, csv readable names, modify the confidence levels of the results
and more. The other functions exist exlcusively to easily acquire the data
required to run HPAStainR.

License Artistic-2.0

Encoding UTF-8

LazyData FALSE

Depends R (>= 4.0.0), dplyr, tidyr

Imports utils, stats, scales, stringr, tibble, shiny, data.table

Suggests knitr, qpdf, testthat

RoxygenNote 7.1.1

SystemRequirements 4GB of RAM

biocViews GeneExpression, GeneSetEnrichment

BugReports https://github.com/tnieuwe/HPAstainR

VignetteBuilder knitr

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

git_branch RELEASE_3_12

git_last_commit ead97ed

git_last_commit_date 2021-02-03

Date/Publication 2021-03-29

Author Tim O. Nieuwenhuis [aut, cre] (<https://orcid.org/0000-0002-1995-3317>)

Maintainer Tim O. Nieuwenhuis <tnieuwe1@jhmi.edu>

R topics documented:

1

HPAStainR	*HPAStainR*

Description

Uses a protein/gene list to query Human Protein Atlas (HPA) staining data.

Usage

```
HPAStainR(
  gene_list,
  hpa_dat,
  cancer_dat = data.frame(),
  cancer_analysis = c("normal", "cancer", "both"),
  tissue_level = TRUE,
  stringency = c("normal", "high", "low"),
  scale_abundance = TRUE,
  round_to = 2,
  csv_names = TRUE,
  stained_gene_data = TRUE,
  tested_protein_column = TRUE,
  percent_or_count = c("percent", "count", "both"),
  drop_na_row = FALSE,
  adjusted_pvals = TRUE
)
```

Arguments

gene_list   A list of proteins or genes that you want to query the HPA staining data with.

hpa_dat   The data frame of normal HPA staining data data, required to run HPAStainR.

cancer_dat   The data frame of pathologic HPA staining data, required to run HPAStainr.

cancer_analysis

                  A character string indicating inclusion of cancer data in the result, must be one
                  of 'normal' (default), 'cancer', or 'both'.

tissue_level   A boolean that determines whether tissue level data for the cell types are in-
                  cluded. Default is TRUE

stringency   A character string indicating how stringent the confidence level of the staining
                  findings have to be. Must be 'normal' (default), 'high', or 'low'.

scale_abundance

                  A boolean that determines whether you scale Staining Score based on the size
                  of the gene list. Default is TRUE.

round_to   A numeric that determines how many decimals in numeric outputs are desired.
                  Default 2.

csv_names   A Boolean determining if you want names suited for a csv file/pipeline, or for
                  presentation. Default is TRUE giving csv names.

stained_gene_data

                  A boolean determining if there is a list of which proteins stained, TRUE is de-
                  fault.

tested_protein_column

> A boolean determining if there is a column listing which proteins were tested, TRUE is default.

percent_or_count

> A character string determining if percent of proteins stained, count of proteins stained, or both are shown for high, medium, and low staining. Must be 'percent' (default), 'count', or 'both'.

drop_na_row   A boolean that determines if cell types with no proteins tested are kept or dropped, default is FALSE.

adjusted_pvals A boolean indicating if you want the p-values corrected for multiple testing. Default is TRUE.

Value

A tibble containing the results of HPAStainR.

Details

Calculation of the staining score below:

$$(\frac{h \times 100}{t}) + (\frac{m \times 50}{t}) + (\frac{l \times 25}{t})$$

Examples

```
## Below will give you the results found on the shiny app website
## This example also uses HPA_data_downloader output as an example
HPA_data <- HPA_data_downloader(tissue_type = 'both', save_file = FALSE)
HPA_out <- HPAStainR(c('PRSS1', 'PNLIP', 'CELA3A', 'PRL'),
HPA_data$hpa_dat,
HPA_data$cancer_dat,
'both')
```

HPA_data_downloader   *HPA_data_downloader*

Description

Used to download required data for HPAStainR

Usage

```
HPA_data_downloader(
  tissue_type = c("both", "normal", "cancer"),
  save_file = TRUE,
  save_location = ""
)
```

Arguments

tissue_type   A character string that determines which HPA data you want to download from
                  the website. Has to be both' (default), 'normal', or 'cancer'.

save_file   A boolean determining if you want the HPA data downloaded permanently
                  or temporarily. Default is TRUE, meaning the file will be saved in the given
                  'save_location', default being the current working directory.

save_location   A character string indicating where you want the files to be saved if you are
                  saving them. If the file(s) already exists in that location, those will be loaded
                  instead of redownloading the files.

Value

List of dataframes or dataframe depending on tissue_type arguement. If tissue_type == 'both' it
will be a list of dataframes.

Examples

```
HPA_data <- HPA_data_downloader(tissue_type = 'both', save_file = FALSE)
## Access normal data
HPA_data$hpa_dat
## Access cancer data
HPA_data$cancer_dat


## Download only the normal tissue data
HPA_normal_data <- HPA_data_downloader('normal', save_file = FALSE)
```

HPA_summary_maker   *HPA_summary_maker*

Description

Used to generate a summary file used in the second tab of the Shiny app version of HPAStainR

Usage

```
HPA_summary_maker(hpa_dat)
```

Arguments

hpa_dat   The dataframe of normal tissue data downloaded by HPA_data_downloader()

Value

A dataframe summarizing the amount of proteins tested to detected, used for the shiny app.

Examples

```
## Load in data from downloader
HPA_data <- HPA_data_downloader(save_file = FALSE)
## Generate the summarized HPA file
hpa_summary <- HPA_summary_maker(HPA_data$hpa_dat)
```

shiny_HPAStainR *shiny_HPAStainR*

Description

Runs HPAStainR as a the shiny app found at https://32tim32.shinyapps.io/HPAStainR/

Usage

```
shiny_HPAStainR(hpa_dat, cancer_dat, cell_type_data = NULL)
```

Arguments

hpa_dat	A required dataframe that has the normal tissue dataframe (see HPA_data_downloader).
cancer_dat	A required dataframe that has the cancer tissue dataframe (see HPA_data_downloader).
cell_type_data	An optional dataframe that comes out of the hpa_summary_maker function, only needed if you want the second tab of HPAStainR, which shows the ratio of tested proteins to stained proteins, to be functional.

Value

A locally ran shiny app

Examples

```
## Load in data from downloader
HPA_data <- HPA_data_downloader(save_file = FALSE)
## Generate the summarized HPA file
hpa_summary <- HPA_summary_maker(HPA_data$hpa_dat)
## Run with summary, commented out so example doesn't run indefinitely
## shiny_HPAStainR(HPA_data$hpa_dat, HPA_data$cancer_dat, hpa_summary)
```

Index