Package 'hpar'

March 30, 2021

Type Package

Title Human Protein Atlas in R

Version 1.32.1

Description The hpar package provides a simple R interface to and
data from the Human Protein Atlas project.

Depends R (>= 3.5.0)

Imports utils

Suggests org.Hs.eg.db, GO.db, knitr, BiocStyle, testthat

VignetteBuilder knitr

License Artistic-2.0

Collate environment.R hpa.R zzz.R

biocViews Proteomics, Homo_sapiens, CellBiology

RoxygenNote 7.1.1

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

git_branch RELEASE_3_12

git_last_commit 68bed75

git_last_commit_date 2020-11-01

Date/Publication 2021-03-29

Author Laurent Gatto [cre, aut] (<https://orcid.org/0000-0002-1520-2268>),
Manon Martin [aut]

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

R topics documented:

1

---

getHpa   *HPA gene query*

Description

Queries one if the HPA data sets with the `id` Ensembl gene identifier. The data set to be used is defined by the `hpadata` argument.

Usage

```
getHpa(id, hpadata = NULL, type = c("data", "details"))
```

Arguments

`id`	A Ensembl gene identifier.
`hpadata`	A `character` with the data set to query. One of available hpar datasets, available by calling [allHparData](or any unambiguous prefix), or `NULL` (default). In the latter case, the default option is used. See [setHparOptions](and [getHparOptions] for more details.
`type`	A `character` defining what type data to return. One of `data` (default, for the data as `data.frame`) or `details` (to open the `id`'s HPA webpage).

Value

A `data.frame` with the information corresponding to the respective `id` genes. If `type` is `details`, then the `dataframe` is returned invisibly and a web page is opened with [browseURL].

Author(s)

Laurent Gatto

Examples

```
id <- "ENSG00000000003"
## Define 'hpadata' data manually
getHpa(id, hpadata = "hpaSubcellularLoc")
head(getHpa(id, hpadata = "hpaNormalTissue"), )
head(getHpa(id, hpadata = "rnaGeneTissue"))
head(getHpa(id, hpadata = "rnaGeneCellLine"))
head(getHpa(id, hpadata = "hpaCancer"))
## Sets default to "SubcellularLoc"
setHparOptions(hpadata = "hpaSubcellularLoc")
getHpa(id)       ## now uses "hpaSubcellularLoc"
setHparOptions() ## reset to "hpaNormalTissue"
head(getHpa(id))
## multiple ids
getHpa(id = c("ENSG00000000003", "ENSG00000000005"),
       hpadata = "hpaSubcellularLoc")
## Not run:
## opens a browser with http://www.proteinatlas.org/ENSG00000163435
getHpa("ENSG00000163435", type = "details")

## End(Not run)
```

---

getHpaDate   *Obtain HPA release information*

Description

Obtain release information about the HPA data, including version, release data and Ensembl version the data is based on. Detailed release information is accessed on-line using `browseURL`.

Usage

```
getHpaDate()
getHpaVersion()
getHpaEnsembl()
```

Value

A `character`.

Author(s)

Laurent Gatto <lg390@cam.ac.uk>

References

HPA release history: http://www.proteinatlas.org/about/releases

Examples

```
getHpaVersion()
getHpaDate()
getHpaEnsembl()
```

---

getHparOptions   *Returns the* hpar *option.*

Description

Returns the `hapr` options. Currently, there is only one, hpadata. See `setHparOptions` for details and examples.

Usage

```
getHparOptions()
```

Value

A `list` with the hpar option(s).

Author(s)

Laurent Gatto

---

hpar   *Human Protein Atlas in R*

Description

This package provides a simple interface to the Human Protein Atlas. From the Human Protein Atlas Project page: The Swedish Human Protein Atlas project, funded by the Knut and Alice Wallenberg Foundation, has been set up to allow for a systematic exploration of the human proteome using Antibody-Based Proteomics. This is accomplished by combining high-throughput generation of affinity-purified antibodies with protein profiling in a multitude of tissues and cells assembled in tissue microarrays. Confocal microscopy analysis using human cell lines is performed for more detailed protein localization. The program hosts the Human Protein Atlas portal with expression profiles of human proteins in tissues and cells.

Details

Several flat files are distributed by the HPA project and available within the package as data.frames, other datasets are available through a search query on the HPA website. The description below is taken from the HPA site:

hpaNormalTissue Normal tissue data: Expression profiles for proteins in human tissues based on immunohistochemisty using tissue micro arrays. The tab-separated file includes Ensembl gene identifier ("Gene"), tissue name ("Tissue"), annotated cell type ("Cell type"), expression value ("Level"), and the gene reliability of the expression value ("Reliability").

hpaNormalTissue16.1 Same as above, for version 16.1

hpaCancer Pathology data: Staining profiles for proteins in human tumor tissue based on immunohistochemisty using tissue micro arrays and log-rank P value for Kaplan-Meier analysis of correlation between mRNA expression level and patient survival. The tab-separated file includes Ensembl gene identifier ("Gene"), gene name ("Gene name"), tumor name ("Cancer"), the number of patients annotated for different staining levels ("High", "Medium", "Low" & "Not detected") and log-rank p values for patient survival and mRNA correlation ("prognostic - favourable", "unprognostic - favourable", "prognostic - unfavourable", "unprognostic - unfavourable").

hpaCancer16.1 Same as above, for version 16.1

rnaGeneTissue RNA HPA tissue gene data: Transcript expression levels summarized per gene in 37 tissues based on RNA-seq. The tab-separated file includes Ensembl gene identifier ("Gene"), analysed sample ("Tissue"), transcripts per million ("TPM"), protein-transcripts per million ("pTPM") and normalized expression ("NX").

rnaGeneCellLine RNA HPA cell line gene data: Transcript expression levels summarized per gene in 64 cell lines. The tab-separated file includes Ensembl gene identifier ("Gene"), analysed sample ("Cell line"), transcripts per million ("TPM"), protein-coding transcripts per million ("pTPM") and normalized expression ("NX").

rnaGeneCellLine16.1 Same as above, for version 16.1

hpaSubcellularLoc Subcellular location data: Subcellular location of proteins based on immunofluorescently stained cells. The tab-separated file includes the following columns: Ensembl gene identifier ("Gene"), name of gene ("Gene name"), gene reliability score ("Reliability"), enhanced locations ("Enhanced"), supported locations ("Supported"), Approved locations ("Approved"), uncertain locations ("Uncertain"), locations with single-cell variation in intensity ("Single-cell variation intensity"), locations with spatial single-cell variation

("Single-cell variation spatial"), locations with observed cell cycle dependency (type can be one or more of biological definition, custom data or correlation) ("Cell cycle dependency"), Gene Ontology Cellular Component term identifier ("GO id").

hpaSubcellularLoc16.1 Same as above, for version 16.1

hpaSubcellularLoc14 Same as above, for version 14.

hpaSecretome Secretome data: The human secretome is here defined as all Ensembl genes with at least one predicted secreted transcript according to HPA predictions. The complete information about the HPA Secretome data is given on `https://www.proteinatlas.org/humanproteome/blood/secretome`. This dataset has 230 columns and includes the Ensembl gene identifier ("Gene"). Information about the additionnal variables can be found here by clicking on *Show/hide columns*.

Detailed description for gene entries and images and not included in the package but can be accessed from within the R environment through a web browser while on-line.

The full data sets can be individually loaded using the data function (see example below). Data about individual genes of interest can retrived with the getHpa function.

HPA data usage policy: The use of data and images from this site in publications and presentations is permitted provided that the following conditions are met:

1. The publication and/or presentation are solely for informational and non-commercial purposes.

2. The source of the data and/or image is referred to this site (www.proteinatlas.org) and/or one or more of our publications are cited.

Author(s)

Laurent Gatto <laurent.gatto@uclouvain.be>

References

See the Human Protein Atlas Project page `http://www.proteinatlas.org/` and `http://www.proteinatlas.org/about/download` for more details and documentation.

Uhlen et al (2015). Tissue-based map of the human proteome. Science. 347(6220):1260419.

Uhlen et al (2010). Towards a knowledge-based Human Protein Atlas. Nat Biotechnol. 28(12):1248-50.

Berglund et al (2008). A gene-centric Human Protein Atlas for expression profiles based on antibodies. Mol Cell Proteomics. 7(10):2019-27.

Uhlen et al (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol Cell Proteomics. 4(12):1920-1932.

Ponten et al (2008). The Human Protein Atlas - a tool for pathology. J Pathology. 216(4):387-93.

See Also

getHpaDate for release information. Gene-specific information should be accessed using the getHpa function.

The package vignette can be accessed with vignette("hpar").

Examples

```
allHparData()

data(hpaNormalTissue)
head(hpaNormalTissue)

data(rnaGeneCellLine)
head(rnaGeneCellLine)

data(rnaGeneTissue)
head(rnaGeneTissue)

data(hpaSubcellularLoc)
head(hpaSubcellularLoc)

data(hpaSubcellularLoc14)
head(hpaSubcellularLoc14)

data(hpaCancer)
head(hpaCancer)

data(hpaSecretome)
head(hpaSecretome)
```

---

setHparOptions   *Sets the 'hpar' options*

Description

Function used to set the hpadata hpar option.

Usage

```
setHparOptions(hpadata = allHparData())
```

Arguments

hpadata   A character defining the default data to be queries. Choices are any of available hpar datasets, available by calling [allHparData](#), or any unambiguous prefix.

Value

Returns the set value of the hpar option. Used for its side effect.

Author(s)

Laurent Gatto

Examples

```
oldOpt <- getHparOptions()
oldOpt
setHparOptions("hpaNormalTissue")
getHparOptions()
setHparOptions(oldOpt$hpadata)
```

Index