

# Package ‘HPAanalyze’

May 7, 2025

**Type** Package

**Title** Retrieve and analyze data from the Human Protein Atlas

**Date** 2023-06-02

**Version** 1.26.0

**Description** Provide functions for retrieving, exploratory analyzing and visualizing the Human Protein Atlas data.

**Depends** R (>= 3.5.0)

**License** GPL-3 + file LICENSE

**RoxygenNote** 7.2.1

**Encoding** UTF-8

**Imports** dplyr, openxlsx, ggplot2, tibble, xml2, stats, utils,  
gridExtra

**Suggests** knitr, rmarkdown, markdown, devtools, BiocStyle

**VignetteBuilder** knitr

**LazyData** true

**biocViews** Proteomics, CellBiology, Visualization, Software

**URL** <https://github.com/anhtr/HPAanalyze>

**BugReports** <https://github.com/anhtr/HPAanalyze/issues>

**git\_url** <https://git.bioconductor.org/packages/HPAanalyze>

**git\_branch** RELEASE\_3\_21

**git\_last\_commit** 15cfc31

**git\_last\_commit\_date** 2025-04-15

**Repository** Bioconductor 3.21

**Date/Publication** 2025-05-07

**Author** Anh Nhat Tran [aut, cre]

**Maintainer** Anh Nhat Tran <trannhatanh89@gmail.com>

Contents

|                               |           |
|-------------------------------|-----------|
| hpaDownload . . . . .         | 2         |
| hpaExport . . . . .           | 4         |
| hpaSubset . . . . .           | 5         |
| hpaVis . . . . .              | 6         |
| hpaVisPatho . . . . .         | 8         |
| hpaVisSubcell . . . . .       | 9         |
| hpaVisTissue . . . . .        | 10        |
| hpaXml . . . . .              | 11        |
| hpaXmlAntibody . . . . .      | 12        |
| hpaXmlGet . . . . .           | 13        |
| hpaXmlProtClass . . . . .     | 14        |
| hpaXmlTissueExpr . . . . .    | 15        |
| hpaXmlTissueExprSum . . . . . | 16        |
| hpa_histology_data . . . . .  | 17        |
| <b>Index</b>                  | <b>18</b> |

---

|             |                          |
|-------------|--------------------------|
| hpaDownload | <i>Download datasets</i> |
|-------------|--------------------------|

---

Description

Download the latest version of HPA datasets and import them in R. It is recommended to only download the datasets you need, as some of them may be very big.

Usage

```
hpaDownload(downloadList = "histology", version = "latest")
```

Arguments

- |              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| downloadList | A vector or string indicate which datasets to download. Possible value: <ul style="list-style-type: none"><li>• 'Normal tissue'</li><li>• 'Pathology'</li><li>• 'Subcellular location'</li><li>• 'RNA consensus tissue'</li><li>• 'RNA HPA tissue'</li><li>• 'RNA GTEx tissue'</li><li>• 'RNA FANTOM tissue'</li><li>• 'RNA single cell type'</li><li>• 'RNA single cell type tissue cluster'</li><li>• 'RNA GTEx brain region'</li><li>• 'RNA FANTOM brain region'</li><li>• 'RNA pig brain region'</li></ul> |
|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

- 'RNA pig brain subregion sample'
- 'RNA mouse brain region'
- 'RNA mouse brain subregion sample'
- 'RNA Allen mouse brain region'
- 'RNA HPA immune cell'
- 'RNA HPA immune cell sample'
- 'RNA Monaco immune cell'
- 'RNA Schmiedel immune cell'
- 'RNA HPA blood cell' (version 21.1)
- 'RNA HPA blood cell sample' (version 21.1)
- 'RNA Monaco blood cell' (version 21.1)
- 'RNA Schmiedel blood cell' (version 21.1)
- 'RNA HPA cell line cancer'
- 'RNA HPA cell line'
- 'RNA TCGA cancer sample'
- 'RNA transcript tissue'
- 'RNA transcript GTEx retina'
- 'RNA transcript immune cells'
- 'RNA transcript cell line' (version 21.1)
- 'RNA transcript pig brain'
- 'RNA transcript mouse brain'

You can also use the following shortcuts:

- 'all': download everything
- 'histology': same as c('Normal tissue', 'Pathology', 'Subcellular location')
- 'rna tissue': same as c('RNA consensus tissue', 'RNA HPA tissue', 'RNA GTEx tissue', 'RNA FANTOM tissue')
- 'rna cell type': same as c('RNA single cell type', 'RNA single cell type tissue cluster')
- 'rna brain region': same as c('RNA GTEx brain region', 'RNA FANTOM brain region', 'RNA pig brain region', 'RNA pig brain subregion sample', 'RNA mouse brain region', 'RNA mouse brain subregion sample', 'RNA Allen mouse brain region')
- 'rna immune cell': same as c('RNA HPA immune cell', 'RNA HPA immune cell sample', 'RNA Monaco immune cell', 'RNA Schmiedel immune cell')
- 'rna blood cell': same as c('RNA HPA blood cell', 'RNA HPA blood cell sample', 'RNA Monaco blood cell', 'RNA Schmiedel blood cell')
- 'isoform': same as c('RNA transcript tissue', 'RNA transcript GTEx retina', 'RNA transcript immune cells', 'RNA transcript cell line', 'RNA transcript pig brain', 'RNA transcript mouse brain')

See <https://www.proteinatlas.org/about/download> for more information.

version

A string indicate which version to be downloaded. Possible value:

- 'latest': Download latest version. Certain legacy datasets will be downloaded with the highest version available. Require Internet connection. This is the default option.
- 'example' or 'built-in': Load the built-in histology dataset from 'HPA-analyze' ('hpa\_histology\_data'). Do not require internet connection.

### Value

This function will return a list of tibbles corresponding to requested datasets.

### See Also

[hpaDownload hpa\\_histology\\_data](#)

Other downloadable datasets functions: [hpaExport\(\)](#), [hpaSubset\(\)](#)

### Examples

```
histologyData <- hpaDownload(downloadList='histology', version='example')
# tissueTranscriptData <- hpaDownload('RNA transcript tissue')
```

---

|           |                               |
|-----------|-------------------------------|
| hpaExport | <i>Export the subset data</i> |
|-----------|-------------------------------|

---

### Description

Export the list object generated by `hpaSubset()` into `xlsx` format. Due to the size of some HPA datasets, as well as the limitation of the output format, exporting the full datasets generated by `hpaDownload()` is not recommended. This is a convenient wrapper for 'write.' functions.

### Usage

```
hpaExport(data, fileName, fileType = "xlsx")
```

### Arguments

|                       |                                                                                                      |
|-----------------------|------------------------------------------------------------------------------------------------------|
| <code>data</code>     | Input the list object generated by <code>hpaSubset()</code>                                          |
| <code>fileName</code> | A string indicate the desired output file name. Do not include file extension such as '.xlsx'.       |
| <code>fileType</code> | The format as which the data will be exported. Choose one of these options: 'xlsx', 'csv' and 'tsv'. |

### Value

- 'xlsx': return one `.xlsx` file named '`fileName.xlsx`'. One individual sheet for each dataset in the input list object.
- 'csv': return `.csv` files, one for each dataset in the input list object, named '`fileName_datasetName.csv`'
- 'tsv': return `.tsv` files, one for each dataset in the input list object, named '`fileName_datasetName.tsv`'

**See Also**

Other downloadable datasets functions: [hpaDownload\(\)](#), [hpaSubset\(\)](#)

**Examples**

```
downloadedData <- hpaDownload(downloadList='histology', version='example')
geneList <- c('TP53', 'EGFR')
tissueList <- c('breast', 'cerebellum', 'skin 1')
cancerList <- c('breast cancer', 'glioma', 'melanoma')

subsetData <- hpaSubset(data=downloadedData,
                        targetGene=geneList,
                        targetTissue=tissueList,
                        targetCancer=cancerList)
hpaExport(data=subsetData,
          fileName='TP53_EGFR_in_tissue_cancer.xlsx',
          fileType='xlsx')
```

---

|           |                               |
|-----------|-------------------------------|
| hpaSubset | <i>Subset downloaded data</i> |
|-----------|-------------------------------|

---

**Description**

`hpaSubset()` subsets data by gene name, tissue, cell type, cancer and/or cell line. The input is the list object generated by `hpaDownload()` or as the output of another `hpaSubset()`. Use `hpaListParam()` to see the list of available parameters for a specific list object. This is a convenient wrapper for ‘`lapply/filter`’ and works on any table which contain ‘`gene`’, ‘`tissue`’, ‘`cell_type`’, ‘`cancer`’, and ‘`cell_line`’ columns.

`hpaListParam()` list available variables in downloaded data that can be used as parameters to subset the data via `hpaSubset()`. This function work with the data object generated by `hpaDownload()` or a previous call of `hpaSubset()`. This is a convenient wrapper for ‘`lapply/unique`’ and works on any table which contain ‘`tissue`’, ‘`cell_type`’, ‘`cancer`’, and ‘`cell_line`’ columns.

**Usage**

```
hpaSubset(
  data = NULL,
  targetGene = NULL,
  targetTissue = NULL,
  targetCellType = NULL,
  targetCancer = NULL,
  targetCellLine = NULL
)

hpaListParam(data = NULL)
```

**Arguments**

|                             |                                                                                                                                                                                                                           |
|-----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>data</code>           | Input the list object generated by <code>hpaDownload()</code> or <code>hpaSubset()</code>                                                                                                                                 |
| <code>targetGene</code>     | Vector of strings of HGNC gene symbols. It will be used to subset every dataset in the list object. You can also mix HGNC gene symbols and ensembl ids (start with ENSG) and they will be converted to HGNC gene symbols. |
| <code>targetTissue</code>   | Vector of strings of normal tissues. Will be used to subset the <code>normal_tissue</code> and <code>rna_tissue</code> dataset.                                                                                           |
| <code>targetCellType</code> | Vector of strings of normal cell types. Will be used to subset the <code>normal_tissue</code> dataset.                                                                                                                    |
| <code>targetCancer</code>   | Vector of strings of cancer types. Will be used to subset the pathology dataset.                                                                                                                                          |
| <code>targetCellLine</code> | Vector of strings of cell lines. Will be used to subset the <code>rna_cell_line</code> dataset.                                                                                                                           |

**Value**

`hpaSubset` will return a list of tibbles as the result of subsetting, depending on the input data.

The output of `hpaListParam()` is a list of vectors containing all subset parameter for the downloaded data.

**See Also**

Other downloadable datasets functions: [hpaDownload\(\)](#), [hpaExport\(\)](#)

**Examples**

```
downloadedData <- hpaDownload(downloadList='histology', version='example')
geneList <- c('TP53', 'EGFR')
tissueList <- c('breast', 'cerebellum', 'skin 1')
cancerList <- c('breast cancer', 'glioma', 'melanoma')

subsetData <- hpaSubset(data=downloadedData,
                        targetGene=geneList,
                        targetTissue=tissueList,
                        targetCancer=cancerList)

downloadedData <- hpaDownload(downloadList='histology', version='example')
params <- hpaListParam(data=downloadedData)
params$normal_tissue
```

## Description

This function is an universal visualization function that allow calling other hpaVis functions via a single function call. By default, this function will use the dataset bundled with HPAanalyze, and provide a grid of all available plots. The types of plots in the output can be specified via the `visType` argument. If only one plot type is specified, this function will return the exact same output as the specific hpaVis function used to create the plot.

## Usage

```
hpaVis(
  data = NULL,
  targetGene = NULL,
  targetTissue = NULL,
  targetCellType = NULL,
  targetCancer = NULL,
  visType = c("Tissue", "Patho", "Subcell"),
  color = c("#FCFDBF", "#FE9F6D", "#DE4968", "#8C2981"),
  customTheme = FALSE,
  ...
)
```

## Arguments

|                             |                                                                                                                                                                                                                                                                                              |
|-----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>data</code>           | Input the list object generated by <code>hpa_download()</code> or <code>hpa_subset()</code> . By default this function use the example dataset bundled with HPAanalyze.                                                                                                                      |
| <code>targetGene</code>     | Vector of strings of HGNC gene symbols. By default it is set to <code>c('TP53', 'EGFR', 'CD44', 'PTEN', 'IDH1')</code> . You can also mix HGNC gene symbols and ensembl ids (start with ENSG) and they will be converted to HGNC gene symbols.                                               |
| <code>targetTissue</code>   | Vector of strings of normal tissue names. By default it is set to <code>"breast"</code> .                                                                                                                                                                                                    |
| <code>targetCellType</code> | Vector of strings of normal cell types. By default includes all available cell types in the target tissues.                                                                                                                                                                                  |
| <code>targetCancer</code>   | Vector of strings of normal tissues. By default it is set to <code>"breast cancer"</code> .                                                                                                                                                                                                  |
| <code>visType</code>        | Vector of strings indicating which plots will be generated. Currently available values are <code>"all"</code> , <code>"Tissue"</code> , <code>"Patho"</code> , <code>"Cancer"</code> , <code>"Subcell"</code> .                                                                              |
| <code>color</code>          | Vector of 4 colors used to depict different expression levels.                                                                                                                                                                                                                               |
| <code>customTheme</code>    | Logical argument. If TRUE, the function will return a barebone ggplot2 plot to be customized further.                                                                                                                                                                                        |
| <code>...</code>            | Additional arguments to be passed downstream to other hpaVis functions being called behind the scene. These arguments includes <code>targetTissue</code> , <code>targetCellType</code> , <code>targetCancer</code> . See documentation for individual hpaVis functions for more information. |

## Value

If multiple `visType` is chosen, this function will return multiple graphs in one panel. If only one `visType` is chosen, this function will return a ggplot2 plot object, which can be further modified

if desirable. See help file for each of the hpaVis function for more information about individual graphs.

See Also

[hpaDownload](#), [hpaSubset](#)  
Other visualization functions: [hpaVisPatho\(\)](#), [hpaVisSubcell\(\)](#), [hpaVisTissue\(\)](#)

Examples

hpaVis()

---

|             |                                 |
|-------------|---------------------------------|
| hpaVisPatho | <i>Visualize pathology data</i> |
|-------------|---------------------------------|

---

Description

Visualize the expression of genes of interest in each cancer.

Usage

```
hpaVisPatho(  
  data = NULL,  
  targetGene = NULL,  
  targetCancer = NULL,  
  facetBy = "cancer",  
  color = c("#FCFDBF", "#FE9F6D", "#DE4968", "#8C2981"),  
  customTheme = FALSE  
)
```

Arguments

|              |                                                                                                                                                                                                                                        |
|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| data         | Input the list object generated by <code>hpa_download()</code> or <code>hpa_subset()</code> . Require the pathology dataset. Use HPA histology data (built-in) by default.                                                             |
| targetGene   | Vector of strings of HGNC gene symbols. By default it is set to <code>c('TP53', 'EGFR', 'CD44', 'PTEN')</code> . You can also mix HGNC gene symbols and ensembl ids (start with ENSG) and they will be converted to HGNC gene symbols. |
| targetCancer | Vector of strings of normal tissues. The function will plot all available cancer by default.                                                                                                                                           |
| facetBy      | Determine how multiple graphs would be faceted. Either cancer (default) or gene.                                                                                                                                                       |
| color        | Vector of 4 colors used to depict different expression levels.                                                                                                                                                                         |
| customTheme  | Logical argument. If TRUE, the function will return a barebone ggplot2 plot to be customized further.                                                                                                                                  |

Value

This function will return a ggplot2 plot object, which can be further modified if desirable. The pathology data is visualized as multiple bar graphs, one for each type of cancer. For each bar graph, x axis contains the inquired protein and y axis contains the proportion of patients.

See Also

Other visualization functions: [hpaVisSubcell\(\)](#), [hpaVisTissue\(\)](#), [hpaVis\(\)](#)

Examples

```
data("hpa_histology_data")
geneList <- c('TP53', 'EGFR', 'CD44', 'PTEN', 'IDH1', 'IDH2', 'CYCS')
cancerList <- c('breast cancer', 'glioma', 'melanoma')

## A typical function call
hpaVisPatho(data=hpa_histology_data,
            targetGene=geneList)
```

---

|               |                                            |
|---------------|--------------------------------------------|
| hpaVisSubcell | <i>Visualize subcellular location data</i> |
|---------------|--------------------------------------------|

---

Description

Visualize the the confirmed subcellular locations of genes of interest.

Usage

```
hpaVisSubcell(
  data = NULL,
  targetGene = NULL,
  reliability = c("enhanced", "supported", "approved", "uncertain"),
  color = c("#FCFDBF", "#8C2981"),
  customTheme = FALSE
)
```

Arguments

|             |                                                                                                                                                                                                                                        |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| data        | Input the list object generated by <code>hpa_download()</code> or <code>hpa_subset()</code> . Require the subcellular_location dataset. Use HPA histology data (built-in) by default.                                                  |
| targetGene  | Vector of strings of HGNC gene symbols. By default it is set to <code>c('TP53', 'EGFR', 'CD44', 'PTEN')</code> . You can also mix HGNC gene symbols and ensemnl ids (start with ENSG) and they will be converted to HGNC gene symbols. |
| reliability | Vector of string indicate which reliability scores you want to plot. The default is everything <code>c("enhanced", "supported", "approved", "uncertain")</code> .                                                                      |

|             |                                                                                                       |
|-------------|-------------------------------------------------------------------------------------------------------|
| color       | Vector of 2 colors used to depict if the protein expresses in a location or not.                      |
| customTheme | Logical argument. If TRUE, the function will return a barebone ggplot2 plot to be customized further. |

Value

This function will return a ggplot2 plot object, which can be further modified if desirable. The subcellular location data is visualized as a tile graph, in which the x axis includes the inquired proteins and the y axis contain the subcellular locations.

See Also

Other visualization functions: [hpaVisPatho\(\)](#), [hpaVisTissue\(\)](#), [hpaVis\(\)](#)

Examples

```
data("hpa_histology_data")
geneList <- c('TP53', 'EGFR', 'CD44', 'PTEN', 'IDH1', 'IDH2', 'CYCS')

## A typical function call
hpaVisSubcell(data=hpa_histology_data,
              targetGene=geneList)
```

---

|              |                              |
|--------------|------------------------------|
| hpaVisTissue | <i>Visualize tissue data</i> |
|--------------|------------------------------|

---

Description

Visualize the expression of protein of interest in each target tissue by cell types.

Usage

```
hpaVisTissue(
  data = NULL,
  targetGene = NULL,
  targetTissue = NULL,
  targetCellType = NULL,
  color = c("#FCFDBF", "#FE9F6D", "#DE4968", "#8C2981"),
  customTheme = FALSE
)
```

Arguments

|      |                                                                                                                                                                                             |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| data | Input the list object generated by <code>hpa_download()</code> or <code>hpa_subset()</code> . Require the <code>normal_tissue</code> dataset. Use HPA histology data (built-in) by default. |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                |                                                                                                                                                                                                                                        |
|----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| targetGene     | Vector of strings of HGNC gene symbols. By default it is set to <code>c('TP53', 'EGFR', 'CD44', 'PTEN')</code> . You can also mix HGNC gene symbols and ensembl ids (start with ENSG) and they will be converted to HGNC gene symbols. |
| targetTissue   | Vector of strings of normal tissues. Default to all.                                                                                                                                                                                   |
| targetCellType | Vector of strings of normal cell types. Default to all.                                                                                                                                                                                |
| color          | Vector of 4 colors used to depict different expression levels.                                                                                                                                                                         |
| customTheme    | Logical argument. If TRUE, the function will return a barebone ggplot2 plot to be customized further.                                                                                                                                  |

### Value

This function will return a ggplot2 plot object, which can be further modified if desirable. The tissue data is visualized as a heatmap: x axis contains inquired protein and y axis contains tissue/cells of interest.

### See Also

Other visualization functions: [hpaVisPatho\(\)](#), [hpaVisSubcell\(\)](#), [hpaVis\(\)](#)

### Examples

```
data("hpa_histology_data")
geneList <- c('TP53', 'EGFR', 'CD44', 'PTEN', 'IDH1', 'IDH2', 'CYCS')
tissueList <- c('breast', 'cerebellum', 'skin 1')

## A typical function call
hpaVisTissue(data=hpa_histology_data,
             targetGene=geneList,
             targetTissue=tissueList)
```

---

|        |                                                                                  |
|--------|----------------------------------------------------------------------------------|
| hpaXml | <i>Extract details about an individual protein from XML file in one function</i> |
|--------|----------------------------------------------------------------------------------|

---

### Description

This function is the umbrella function for the hpaXml function family. It take the input of either one Ensembl gene id or a imported XML object resulting from a hpaXmlGet() function call. By default, it will extract all information available for HPAanalyze user from the XML file by calling every hpaXml function and put all results into a list.

### Usage

```
hpaXml(
  inputXml,
  extractType = c("ProtClass", "TissueExprSum", "Antibody", "TissueExpr"),
  ...
)
```

**Arguments**

|             |                                                                                                                                                                                                                        |
|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| inputXml    | Input can be either one Ensembl gene id (start with ENSG) or a imported XML object resulting from a hpaXmlGet() function call. You can also use HGNC gene symbol and it will be converted to ensembl id.               |
| extractType | A vector of strings indicate which information is desired for extraction. By default this function will call all hpaXml functions available. Other options are 'ProtClass', 'TissueExprSum', 'Antibody', 'TissueExpr'. |
| ...         | Additional arguments to be passed downstream to other hpaXml functions being called behind the scene. See help files of other hpaXml functions for more information.                                                   |

**Value**

This function returns a list. Each element of the list is information extracted from the XML file specified using other hpaXml functions. See help file for each XML function for more information.

**See Also**

Other xml functions: [hpaXmlAntibody\(\)](#), [hpaXmlGet\(\)](#), [hpaXmlProtClass\(\)](#), [hpaXmlTissueExprSum\(\)](#), [hpaXmlTissueExpr\(\)](#)

**Examples**

```
hpaXml(inputXml='ENSG00000131979', extractType=c('ProtClass', 'TissueExprSum', 'Antibody'))
```

---

|                |                                     |
|----------------|-------------------------------------|
| hpaXmlAntibody | <i>Extract antibody information</i> |
|----------------|-------------------------------------|

---

**Description**

Extract information about the antibodies used for a specific protein. It is important to note that the data that HPA provides on their website and through xml files are not one-to-one equivalents.

**Usage**

```
hpaXmlAntibody(importedXml)
```

**Arguments**

|             |                                                                |
|-------------|----------------------------------------------------------------|
| importedXml | Input an xml document object resulted from a hpaXmlGet() call. |
|-------------|----------------------------------------------------------------|

**Value**

This function returns a tibble of 4 columns, containing information about the antibodies used in the project for the inquired protein: id, releaseDate, releaseVersion, and RRID.

**See Also**

Other xml functions: [hpaXmlGet\(\)](#), [hpaXmlProtClass\(\)](#), [hpaXmlTissueExprSum\(\)](#), [hpaXmlTissueExpr\(\)](#), [hpaXml\(\)](#)

**Examples**

```
## Not run:
GCH1xml <- hpaXmlGet('ENSG00000131979')
hpaXmlAntibody(GCH1xml)

## End(Not run)
```

---

|           |                                     |
|-----------|-------------------------------------|
| hpaXmlGet | <i>Download and import xml file</i> |
|-----------|-------------------------------------|

---

**Description**

Download and import individual xml file for a specified protein. This function calls `xml2::read_xml()` under the hood. It is important to note that the data that HPA provides on their website and through xml files are not one-to-one equivalents.

**Usage**

```
hpaXmlGet(targetEnsemblId, version = "latest")
```

**Arguments**

|                 |                                                                                                                                                                                                                                                                                                                         |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| targetEnsemblId | A string of one ensembl ID, start with ENSG. For example 'ENSG00000131979'. You can also use HGNC gene symbol and it will be converted to ensembl id.                                                                                                                                                                   |
| version         | A string indicate which version to be downloaded. Possible value: <ul style="list-style-type: none"> <li>'latest': Download latest version.</li> <li>'v?' with '?' is a integer: Download a specific version of the dataset. For example: 'v18' download version 18. Currently support version 13 and above.</li> </ul> |

**Value**

This function return an object of class "xml\_document" "xml\_node" containing the content of the imported XML file. (See documentations for package xml2 for more information.)

**See Also**

Other xml functions: [hpaXmlAntibody\(\)](#), [hpaXmlProtClass\(\)](#), [hpaXmlTissueExprSum\(\)](#), [hpaXmlTissueExpr\(\)](#), [hpaXml\(\)](#)

## Examples

```
## Not run:
GCH1xml <- hpaXmlGet('ENSG00000131979')

## End(Not run)
```

---

|                 |                                |
|-----------------|--------------------------------|
| hpaXmlProtClass | <i>Extract protein classes</i> |
|-----------------|--------------------------------|

---

## Description

Extract protein class information from imported xml document resulted from `hpaXmlGet()`. It is important to note that the data that HPA provides on their website and through xml files are not one-to-one equivalents.

## Usage

```
hpaXmlProtClass(importedXml)
```

## Arguments

`importedXml`      Input an xml document object resulted from a `hpaXmlGet()` call.

## Value

This function return a tibble of 4 columns.

## See Also

Other xml functions: [hpaXmlAntibody\(\)](#), [hpaXmlGet\(\)](#), [hpaXmlTissueExprSum\(\)](#), [hpaXmlTissueExpr\(\)](#), [hpaXml\(\)](#)

## Examples

```
## Not run:
GCH1xml <- hpaXmlGet('ENSG00000131979')
hpaXmlProtClass(GCH1xml)

## End(Not run)
```

---

|                  |                                          |
|------------------|------------------------------------------|
| hpaXmlTissueExpr | <i>Extract tissue expression details</i> |
|------------------|------------------------------------------|

---

## Description

Extract tissue expression information for each sample and url to download images from imported xml document resulted from `hpaXmlGet()`. It is important to note that the data that HPA provides on their website and through xml files are not one-to-one equivalents. For example, xml files usually only provide one of the two histology image for each patient.

## Usage

```
hpaXmlTissueExpr(importedXml)
```

## Arguments

`importedXml`      Input an xml document object resulted from a `hpaXmlGet()` call.

## Value

This function returns a list of tibbles, each for an antibody. Each tibble contains information about all individual samples and their staining. Due to the variation in amount of information available for these samples, the number of columns differs, but the tibble essentially includes: `patientId`, `age`, `sex`, `staining`, `intensity`, `quantity`, `location`, `imageUrl`, `snomedCode`, and `tissueDescription`. The last two items may have more than one column each.

## See Also

Other xml functions: [hpaXmlAntibody\(\)](#), [hpaXmlGet\(\)](#), [hpaXmlProtClass\(\)](#), [hpaXmlTissueExprSum\(\)](#), [hpaXml\(\)](#)

## Examples

```
## Not run:  
GCH1xml <- hpaXmlGet('ENSG00000131979')  
hpaXmlTissueExpr(GCH1xml)  
  
## End(Not run)
```

---

|                     |                                                      |
|---------------------|------------------------------------------------------|
| hpaXmlTissueExprSum | <i>Extract tissue expression and download images</i> |
|---------------------|------------------------------------------------------|

---

## Description

Extract tissue expression information and url to download images from imported xml document resulted from `hpaXmlGet()`. It is important to note that the data that HPA provides on their website and through xml files are not one-to-one equivalents.

## Usage

```
hpaXmlTissueExprSum(importedXml, downloadImg = FALSE)
```

## Arguments

|                          |                                                                                                         |
|--------------------------|---------------------------------------------------------------------------------------------------------|
| <code>importedXml</code> | Input an xml document object resulted from a <code>hpaXmlGet()</code> call.                             |
| <code>downloadImg</code> | Logical argument. The function will download all image from the extracted urls into the working folder. |

## Value

This function return a list consists of a summary string, which is a very brief description of the protein, and a tibble of 2 columns: tissue (name of tissue available) and imageUrl (link to download the perspective image)

## See Also

Other xml functions: [hpaXmlAntibody\(\)](#), [hpaXmlGet\(\)](#), [hpaXmlProtClass\(\)](#), [hpaXmlTissueExpr\(\)](#), [hpaXml\(\)](#)

## Examples

```
## Not run:
GCH1xml <- hpaXmlGet('ENSG00000131979')
hpaXmlTissueExprSum(GCH1xml)

## End(Not run)
```

---

|                    |                              |
|--------------------|------------------------------|
| hpa_histology_data | <i>HPA histology dataset</i> |
|--------------------|------------------------------|

---

**Description**

Dataset downloaded with `hpaDownload('histology', version = 'latest')`. This should be the most updated dataset at the time of generation. Check metadata for more information.

**Usage**

```
hpa_histology_data
```

**Format**

A list of 3 tibbles

**normal\_tissue** Normal tissue IHC data

**pathology** Cancer IHC data

**subcellular\_location** Subcellular location IF data

**See Also**

[hpaDownload](#)

**Examples**

```
# load data
data("hpa_histology_data")

# access data frames
normal_tissue_data <- hpa_histology_data$normal_tissue
cancer_data <- hpa_histology_data$pathology
subcell_location_data <- hpa_histology_data$subcellular_location

# see metadata
hpa_histology_data$metadata
```

# Index

## \* datasets

hpa\_histology\_data, [17](#)

## \* downloadable datasets functions

hpaDownload, [2](#)

hpaExport, [4](#)

hpaSubset, [5](#)

## \* visualization functions

hpaVis, [6](#)

hpaVisPatho, [8](#)

hpaVisSubcell, [9](#)

hpaVisTissue, [10](#)

## \* xml functions

hpaXml, [11](#)

hpaXmlAntibody, [12](#)

hpaXmlGet, [13](#)

hpaXmlProtClass, [14](#)

hpaXmlTissueExpr, [15](#)

hpaXmlTissueExprSum, [16](#)

hpa\_histology\_data, [4](#), [17](#)

hpaDownload, [2](#), [4–6](#), [8](#), [17](#)

hpaExport, [4](#), [4](#), [6](#)

hpaListParam (hpaSubset), [5](#)

hpaSubset, [4](#), [5](#), [5](#), [8](#)

hpaVis, [6](#), [9–11](#)

hpaVisPatho, [8](#), [8](#), [10](#), [11](#)

hpaVisSubcell, [8](#), [9](#), [9](#), [11](#)

hpaVisTissue, [8–10](#), [10](#)

hpaXml, [11](#), [13–16](#)

hpaXmlAntibody, [12](#), [12](#), [13–16](#)

hpaXmlGet, [12](#), [13](#), [13](#), [14–16](#)

hpaXmlProtClass, [12](#), [13](#), [14](#), [15](#), [16](#)

hpaXmlTissueExpr, [12–14](#), [15](#), [16](#)

hpaXmlTissueExprSum, [12–15](#), [16](#)