

Package ‘inSilicoDb’

October 8, 2014

Version 2.0.1

Date 2014-04-10

Title Access to the InSilico Database

Description

Access expert curated and normalized microarray eSet datasets from the InSilico Database.

Author Jonatan Taminau <jtaminau@gmail.com>

Maintainer Quentin De Clerck <qdeclerc@vub.ac.be>, David Steenhoff <davidsteehoff@insilicodb.org>

Depends R (>= 3.0.0), rjson, Biobase, RCurl

Imports

Suggests limma

Collate util.R interface.R basic.R

biocViews Microarray, DataImport

License GPL-2

URL <https://insilicodb.com>

R topics documented:

inSilicoDb-package	2
getAnnotations	2
getCurationInfo	3
getDataset	4
getDatasetInfo	5
getDatasetList	7
getDatasetPlatformList	8
getDatasets	10
getDefaultCuration	11

getInSilicoUserDetails	12
getPlatformList	12
getPlatforms	13
InSilicoLogin	14
InSilicoLogout	14

Index	15
--------------	-----------

inSilicoDb-package	<i>Basic access to the InSilico database.</i>
--------------------	---

Description

This package provides basic access to the InSilico database. For full functionality visit: <https://insilicodb.com>

See Also

[getAnnotations](#) [getCurationInfo](#) [getDataset](#) [getDatasetInfo](#) [getDatasetList](#) [getDatasetPlatformList](#)
[getDatasets](#) [getDefaultCuration](#) [getInSilicoUserDetails](#) [getPlatforms](#) [InSilicoLogin](#)
[InSilicoLogout](#)

getAnnotations	<i>Get annotation information</i>
----------------	-----------------------------------

Description

Get annotation information for a given dataset identifier in Bioconductors AnnotatedDataFrame format.

Usage

```
getAnnotations(dataset, ...);
```

Arguments

dataset	Valid dataset identifier: the GEO or InSilico DB accession code.
...	Optional arguments can be passed to the function for more specific requests:
platform	The platform of the chip on which a measurement was done. A platform is identified by its GEO accession code.
curation	Specific curation used to annotate a dataset. A curation is identified by a InSilico DB Curation identifier. You can annotate datasets yourself at https://insilicodb.com/app/browse If not defined, the preferred curation of specified dataset will be used.

Value

A Bioconductors AnnotatedDataFrame.

Examples

```
annot <- getAnnotations("GSE4635");
pData(annot);
##           Anatomical Site  Smoker
## GSM15729      bronchus      never
## GSM104072     bronchus      current
## ...
```

<code>getCurationInfo</code>	<i>Get the available curations for a dataset</i>
------------------------------	--

Description

Get the information about different curations for a given dataset.

Usage

```
getCurationInfo(dataset);
```

Arguments

`dataset` Valid dataset identifier: the GEO or InSilico DB accession code.

Value

Messages describing the curations are printed.

Examples

```
getCurationInfo("GSE4635");
## INSILICODB: 5 curations found.
##
## INSILICODB: =====
## INSILICODB: curation id: 14926 (preferred)
## INSILICODB: =====
## INSILICODB: curator:   Virginie de Schaetzen
## INSILICODB: date:     2011-12-13
## INSILICODB: keywords: Age, Sex, Anatomical Site,
##                      Cell Type, Smoker, Ethnicity,
##                      Cigarette Consumption (pack/years),
##                      platform
## INSILICODB:
## ...
```

getDataset

Get dataset in ExpressionSet format

Description

Get dataset for a given dataset identifier and platform identifier in a specific format.

Usage

```
getDataset(dataset, platform, ...);
```

Arguments

dataset	Valid dataset identifier: the GEO or InSilico DB accession code.
platform	The platform of the chip on which a measurement was done. A platform is identified by its GEO accession code.
...	Optional arguments can be passed to the function for more specific requests: format = "CURESET" File format in which the dataset should be returned. Possible values: <ul style="list-style-type: none"> • CURESET : Curated R/Bioconductor ExpressionSet. • ESET : R/Bioconductor ExpressionSet without curation. features = "PROBE" Type of features in which data is represented. Two possible values: "GENE" or "PROBE". By default an expression matrix containing probes is returned. norm = "ORIGINAL" Type of normalization to use when pre-processing a dataset. Possible normalizations: "ORIGINAL", "FRMA"[1], "SCAN"[2] or "UPC"[3]. curation Specific curation used to annotate a dataset. A curation is identified by a InSilico DB Curation identifier. You can annotate datasets yourself at https://insilicodb.com/app/browse If not defined, the preferred curation of specified dataset will be used. timeout = 120 Time (in seconds) to wait for the dataset, if it has to be generated.

Value

A Bioconductors ExpressionSet.

References

- [1] MN. McCall, BM. Bolstad, and RA. Irizarry, Frozen Robust Multi-Array Analysis (fRMA), *Biostatistics*, vol. 11, no. 2, pp. 242-253, 2008.
- [2] SR. Piccolo, Y. Sun, JD. Campbell, ME. Lenburg, AH. Bild, and WE. Johnson, A single-sample microarray normalization method to facilitate personalized-medicine workflows, *Genomics*, vol. 100, no. 6, pp. 337-344, 2012.

[3] SR. Piccolo, MR. Withers, OE. Francis, AH. Bild, and WE. Johnson, Multi-platform single-sample estimates of transcriptional activation, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 44, pp. 17778-17783, 2013.

See Also

[getDatasetInfo](#)

Examples

```
dataset.id <- "GSE4635";
platform.id <- getPlatforms(dataset.id);
eset <- getDataset(dataset.id, platform.id[[1]],
                  norm = "ORIGINAL", features = "PROBE");
print(dim(eset));
## Features Samples
##    22215      8
eset <- getDataset( dataset.id, platform.id[[1]],
                  norm = "ORIGINAL", features = "GENE");
print(dim(eset));
## Features Samples
##    12698      8
```

getDatasetInfo	<i>Get information about a dataset</i>
----------------	--

Description

Get information about the availability and the default parameters of a dataset.

Usage

```
getDatasetInfo(dataset, platform, ...);
```

Arguments

dataset	Valid dataset identifier: the GEO or InSilico DB accession code.
platform	The platform of the chip on which a measurement was done. A platform is identified by its GEO accession code.
...	Optional arguments can be passed to the function for more specific requests:
	format = "CURESET" File format in which the dataset should be returned. Possible values:
	<ul style="list-style-type: none"> • CURESET : Curated R/Bioconductor ExpressionSet. • ESET : R/Bioconductor ExpressionSet without curation.
	features = "PROBE" Type of features in which data is represented. Two possible values: "GENE" or "PROBE". By default an expression matrix containing probes is returned.

norm = "ORIGINAL" Type of normalization to use when pre-processing a dataset. Possible normalizations: "ORIGINAL", "FRMA"[1], "SCAN"[2] or "UPC"[3].

curation Specific curation used to annotate a dataset. A curation is identified by a InSilico DB Curation identifier. You can annotate datasets yourself at <https://insilicodb.com/app/browse> If not defined, the preferred curation of specified dataset will be used.

Details

The function `getDatasetInfo` has two functionalities.

Its first functionality is to return the default parameters of the given dataset if it is called with `dataset` and `platform` as only arguments.

Its second functionality is to check the availability of a dataset when optional arguments are specified. Do not forget to surround the function call by a `tryCatch` function when checking for availability.

References

- [1] MN. McCall, BM. Bolstad, and RA. Irizarry, Frozen Robust Multi-Array Analysis (fRMA), *Biostatistics*, vol. 11, no. 2, pp. 242-253 , 2008.
- [2] SR. Piccolo, Y. Sun, JD. Campbell, ME. Lenburg, AH. Bild, and WE. Johnson, A single-sample microarray normalization method to facilitate personalized-medicine workflows, *Genomics*, vol. 100, no. 6, pp. 337-344, 2012.
- [3] SR. Piccolo, MR. Withers, OE. Francis, AH. Bild, and WE. Johnson, Multi-platform single-sample estimates of transcriptional activation, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 44, pp. 17778-17783, 2013.

See Also

Is automatically called before `getDataset`.

Examples

```
data.id <- "GSE781";
plt.id <- "GPL96";

## return the default parameters when calling getDataset
eset <- getDatasetInfo( data.id, plt.id);
eset$norm;
## [1] "FRMA"

plt.id <- "GPL97";

## check the availability of the normalizations
norms <- c("FRMA", "ORIGINAL")
output <- sapply(norms, function(n) {
  tryCatch({
    eset <- getDatasetInfo(data.id, plt.id, norm = n);
    eset$norm;
  })
})
```

```

    }, error = function(e) {
      message(e);
    });
  });
  output$FRMA
  ## NULL
  output$ORIGINAL
  ## [1] "ORIGINAL"

```

getDatasetList

Get list of all available dataset identifiers

Description

Get list of all available dataset identifiers that satisfy all specified constraints (see Arguments).

Usage

```
getDatasetList(...);
```

Arguments

... Optional arguments can be passed to the function for more specific requests:

platform The platform of the chip on which a measurement was done. A platform is identified by its GEO accession code. If specified, only datasets on the given platform are returned.

norm Type of normalization to use when pre-processing a dataset. Possible values: "ORIGINAL", "FRMA"[1], "SCAN"[2] or "UPC"[3]. If specified, only datasets for which the given preprocessed data is available are returned.

query If specified, only datasets for which the query keyword appears in either the title or description of the dataset. typical example queries are 'Thyroid Cancer', 'Primary vs Metastasis', 'p53', etc.

curated If curated = TRUE is specified, only manually curated datasets will be returned.

Value

A list of dataset identifiers.

References

- [1] MN. McCall, BM. Bolstad, and RA. Irizarry, Frozen Robust Multi-Array Analysis (fRMA), *Biostatistics*, vol. 11, no. 2, pp. 242-253 , 2008.
- [2] SR. Piccolo, Y. Sun, JD. Campbell, ME. Lenburg, AH. Bild, and WE. Johnson, A single-sample microarray normalization method to facilitate personalized-medicine workflows, *Genomics*, vol. 100, no. 6, pp. 337-344, 2012.

[3] SR. Piccolo, MR. Withers, OE. Francis, AH. Bild, and WE. Johnson, Multi-platform single-sample estimates of transcriptional activation, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 44, pp. 17778-17783, 2013.

See Also

Use `getDatasetPlatformList` for the same functionality with a list of (dataset, platform) tuples as result.

Examples

```
length(getDatasetList());
## 7167
length(getDatasetList(platform = "GPL570"));
## 2717
length(getDatasetList(platform = "GPL570", norm = "FRMA"));
## 2406
length(getDatasetList(platform = "GPL570", norm = "FRMA",
                      query = "Thyroid Cancer"));
## 5
getDatasetList(platform = "GPL570", norm = "FRMA",
               query = "Thyroid Cancer");
## [1] "GSE6004" "GSE7307" "GSE32161" "GSE29265" "GSE2109"
```

getDatasetPlatformList

Get a list of (dataset, platform) tuples

Description

Get list of all available dataset and platform tuples that satisfy all specified constraints (see Arguments).

Usage

```
getDatasetPlatformList(...);
```

Arguments

... Optional arguments can be passed to the function for more specific requests:

- platform The platform of the chip on which a measurement was done. A platform is identified by its GEO accession code. If specified, only datasets on the given platform are returned.
- norm Type of normalization to use when pre-processing a dataset. Possible values: "ORIGINAL", "FRMA"[1], "SCAN"[2] or "UPC"[3]. If specified, only datasets for which the given preprocessed data is available are returned.

query If specified, only datasets for which the query keyword appears in either the title or description of the dataset. typical example queries are 'Thyroid Cancer', 'Primary vs Metastasis', 'p53', etc.

curated If curated = TRUE is specified, only manually curated datasets will be returned.

Value

A list of dataset and platform tuples.

References

[1] MN. McCall, BM. Bolstad, and RA. Irizarry, Frozen Robust Multi-Array Analysis (fRMA), *Biostatistics*, vol. 11, no. 2, pp. 242-253 , 2008.

[2] SR. Piccolo, Y. Sun, JD. Campbell, ME. Lenburg, AH. Bild, and WE. Johnson, A single-sample microarray normalization method to facilitate personalized-medicine workflows, *Genomics*, vol. 100, no. 6, pp. 337-344, 2012.

[3] SR. Piccolo, MR. Withers, OE. Francis, AH. Bild, and WE. Johnson, Multi-platform single-sample estimates of transcriptional activation, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 44, pp. 17778-17783, 2013.

See Also

Use [getDatasetList](#) for the same functionality with a list of dataset identifiers as result.

Examples

```
length(getDatasetPlatformList());
## 7167
length(getDatasetPlatformList(platform = "GPL570"));
## 2717
length(getDatasetPlatformList(platform = "GPL570", norm = "FRMA"));
## 2406
length(getDatasetPlatformList(platform = "GPL570", norm = "FRMA",
                             query = "Thyroid Cancer"));
## 5
getDatasetPlatformList(platform = "GPL570", norm = "FRMA",
                      query = "Thyroid Cancer");
## [[1]]
## [1] "GSE6004" "GPL570"
##
## [[2]]
## [1] "GSE7307" "GPL570"
##
## [[3]]
## [1] "GSE32161" "GPL570"
##
## [[4]]
## [1] "GSE29265" "GPL570"
##
## [[5]]
```

```
## [1] "GSE2109" "GPL570"
```

```
getDatasets
```

```
Get datasets in ExpressionSet format
```

Description

Get the datasets for every platform for the given dataset identifier.

Usage

```
getDatasets(dataset, ...);
```

Arguments

dataset	Valid dataset identifier: the GEO or InSilico DB accession code.
...	Optional arguments can be passed to the function for more specific requests:
	format = "CURESET" File format in which the dataset should be returned. Possible values:
	<ul style="list-style-type: none"> • CURESET : Curated R/Bioconductor ExpressionSet. • ESET : R/Bioconductor ExpressionSet without curation.
	features = "PROBE" Type of features in which data is represented. Two possible values: "GENE" or "PROBE". By default an expression matrix containing probes is returned
	norm = "ORIGINAL" Type of normalization to use when pre-processing a dataset. Possible normalizations: "ORIGINAL", "FRMA"[1], "SCAN"[2] or "UPC"[3].
	curation Specific curation used to annotate a dataset. A curation is identified by a InSilico DB Curation identifier. You can annotate datasets yourself at https://insilicodb.com/app/browse If not defined, the default curation of specified dataset will be used.
	timeout = 120 Time (in seconds) to wait for the dataset, if it has to be generated.

Value

A list with a Bioconductors ExpressionSet for every platform.

References

- [1] MN. McCall, BM. Bolstad, and RA. Irizarry, Frozen Robust Multi-Array Analysis (fRMA), *Biostatistics*, vol. 11, no. 2, pp. 242-253 , 2008.
- [2] SR. Piccolo, Y. Sun, JD. Campbell, ME. Lenburg, AH. Bild, and WE. Johnson, A single-sample microarray normalization method to facilitate personalized-medicine workflows, *Genomics*, vol. 100, no. 6, pp. 337-344, 2012.
- [3] SR. Piccolo, MR. Withers, OE. Francis, AH. Bild, and WE. Johnson, Multi-platform single-sample estimates of transcriptional activation, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 44, pp. 17778-17783, 2013.

Examples

```
dataset.id <- "GSE1456"
getPlatforms(dataset.id);
## [1] "GPL96" "GPL97"
esets <- getDatasets(dataset.id, norm = "ORIGINAL");
sapply(esets, annotation)
##      GPL96      GPL97
## "hgu133a" "hgu133b"
```

getDefaultCuration *Default curation of a dataset*

Description

Get the default curation for a given dataset.

Usage

```
getDefaultCuration(dataset);
```

Arguments

dataset Valid dataset identifier: the GEO or InSilico DB accession code.

Value

List containing curation information: curator id, date and curator name.

Examples

```
getDefaultCuration("GSE4635");
## $id
## [1] "14926"
## $date
## [1] "2011-12-13 15:31:44"
## $curator
## [1] "Virginie de Schaetzen"
```

`getInSilicoUserDetails`*Get the details of your InSilico DB account*

Description

Get your user details when logged in.

Usage

```
getInSilicoUserDetails();
```

See Also

See [InSilicoLogin](#), [InSilicoLogout](#) for more information about logging in and out.

`getPlatformList`*Get list of platform identifiers*

Description

Get the list of all platform identifiers supported by the package.

Usage

```
getPlatformList(...);
```

Arguments

... Optional arguments can be passed to the function for more specific requests:

- `norm` Type of normalization to use when pre-processing a dataset. Possible values: "ORIGINAL", "FRMA"[1], "SCAN"[2] or "UPC"[3]. If specified, only datasets for which the given preprocessed data is available are returned.

Value

A list of platform identifiers.

References

- [1] MN. McCall, BM. Bolstad, and RA. Irizarry, Frozen Robust Multi-Array Analysis (fRMA), *Biostatistics*, vol. 11, no. 2, pp. 242-253 , 2008.
- [2] SR. Piccolo, Y. Sun, JD. Campbell, ME. Lenburg, AH. Bild, and WE. Johnson, A single-sample microarray normalization method to facilitate personalized-medicine workflows, *Genomics*, vol. 100, no. 6, pp. 337-344, 2012.
- [3] SR. Piccolo, MR. Withers, OE. Francis, AH. Bild, and WE. Johnson, Multi-platform single-sample estimates of transcriptional activation, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 44, pp. 17778-17783, 2013.

Examples

```
length(getPlatformList());  
## 14  
length(getPlatformList(norm="ORIGINAL"));  
## 14  
length(getPlatformList(norm="FRMA"));  
## 6
```

getPlatforms	<i>Get platforms for a dataset</i>
--------------	------------------------------------

Description

Get all platforms that are used in a given dataset.

Usage

```
getPlatforms(dataset);
```

Arguments

dataset Valid dataset identifier: the GEO or InSilico DB accession code.

Value

List of valid platform identifiers.

Examples

```
getPlatforms("GSE4635");  
## [1] "GPL96"  
getPlatforms("GSE781");  
## [1] "GPL96" "GPL97"
```

InSilicoLogin	<i>Login into the InSilico DB</i>
---------------	-----------------------------------

Description

Login into the InSilico DB enables the user to access his private data. The password has to be in MD5 hash format.

Usage

```
InSilicoLogin(login, password);
```

Arguments

login	Username or email address.
password	MD5 hash of your password.

See Also

[InSilicoLogout](#), [getInSilicoUserDetails](#)

InSilicoLogout	<i>Logout of InSilico DB</i>
----------------	------------------------------

Description

Logout of InSilico DB.

Usage

```
InSilicoLogout();
```

See Also

[InSilicoLogin](#), [getInSilicoUserDetails](#)

Index

[getAnnotations](#), [2](#), [2](#)
[getCurationInfo](#), [2](#), [3](#)
[getDataset](#), [2](#), [4](#), [6](#)
[getDatasetInfo](#), [2](#), [5](#), [5](#)
[getDatasetList](#), [2](#), [7](#), [9](#)
[getDatasetPlatformList](#), [2](#), [8](#), [8](#)
[getDatasets](#), [2](#), [10](#)
[getDefaultCuration](#), [2](#), [11](#)
[getInSilicoUserDetails](#), [2](#), [12](#), [14](#)
[getPlatformList](#), [12](#)
[getPlatforms](#), [2](#), [13](#)

[inSilicoDb \(inSilicoDb-package\)](#), [2](#)
[inSilicoDb-package](#), [2](#)
[InSilicoLogin](#), [2](#), [12](#), [14](#), [14](#)
[InSilicoLogout](#), [2](#), [12](#), [14](#), [14](#)