Package ‘inSilicoDb’

October 9, 2013

Version 1.8.0

Date 2011-05-24

Title Access to the InSilico Database

Description Access Human Affymetrix expert curated Gene Expression Omnibus (GEO) datasets from the InSilico Database.

Author Jonatan Taminau `<jtaminau@vub.ac.be>`

Maintainer Jonatan Taminau `<jtaminau@vub.ac.be>`, David Steenhoff `<davidsteenhoff@insilicodb.org>`

Depends R (>= 2.11.0), rjson, Biobase

Imports RCurl

Suggests limma

Collate zzz.R util.R basic.R interface.R

biocViews Microarray, DataImport

License GPL-2

URL https://insilicodb.org

R topics documented:

`inSilicoDb-package` *Basic access to the inSilico Database.*

Description

This package provides basic access to the inSilico database. For full functionality visit: http://insilico.ulb.ac.be

See Also

`getPlatforms` `getDataset` `getDatasets` `getAnnotations` `getDatasetList`

`getAnnotations` *Get annotation information*

Description

Get annotation information for a given series GSE and platform GPL identifier in Bioconductors AnnotatedDataFrame format

Usage

```
getAnnotations(gse, gpl);
```

Arguments

`gse`	Valid series GSE identifier.
`gpl`	Valid platform GPL identifier.

Value

A Bioconductors AnnotatedDataFrame

Examples

```
## Not run:
  gse = "GSE4635";
  platforms = getPlatforms(gse);
  annot = getAnnotations("GSE4635", "GPL96");
  pData(annot);
##              Anatomical Site  Smoker
## GSM15729     bronchus         never
## GSM104072    bronchus         current
## ...

## End(Not run)
```

`getDataset` *Get dataset in ExpressionSet format*

Description

Get dataset for a given series GSE and platform GPL identifier in Bioconductors ExpressionSet format

Usage

```
getDataset(gse, gpl, norm="ORIGINAL", genes=FALSE);
```

Arguments

`gse`	Valid series GSE identifier.
`gpl`	Valid platform GPL identifier.
`norm`	Required preprocessing method. Currently available options are ORIGINAL and FRMA. The ORIGINAL normalization is how the original authors submitted it to Gene Expression Omnibus (GEO). This data is always available and is therefore the default option. The availability of FRMA normalization depends on the availability of the CEL files by the original authors.
`genes`	By default a gene expression matrix containing probes is returned. Precomputed conversions using genes instead of probes are also available if this argument is set to TRUE. This conversion was made using the `nsFilter` function from Bioconductors `genefilter` package.

Value

A Bioconductors ExpressionSet

Examples

```
## Not run: 
  gse = "GSE4635";
  platforms = getPlatforms(gse);
  eset = getDataset(gse,platforms[[1]],norm="ORIGINAL",genes=FALSE);
  print(dim(eset));
  ## Features  Samples
  ##    22283        8
  eset = getDataset(gse,platforms[[1]],norm="ORIGINAL",genes=TRUE);
  print(dim(eset));
  ## Features  Samples
  ##    12679        8

## End(Not run)
```

`getDatasetList` *Get list of all available series GSE identifiers*

Description

Get list of all available series GSE identifiers that satisfy platform, normalization and query constraints (see Arguments)

Usage

```
getDatasetList(gpl="", norm="", query="");
```

Arguments

`gpl`	Valid platform GPL identifier. If specified, only datasets on the given platform are returned.
`norm`	Valid preprocessing method. Currently available options are 'ORIGINAL' and 'FRMA'. 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.

Value

A list of GSE identifiers

Examples

```
## Not run:
  length(getDatasetList());
  ## 3173
  length(getDatasetList(gpl="GPL570"));
  ## 1930
  length(getDatasetList(gpl="GPL570", norm="FRMA"));
  ## 1381
  length(getDatasetList(gpl="GPL570", norm="FRMA", query="Thyroid Cancer"));
  ## 8

## End(Not run)
```

`getDatasets` *Get datasets in ExpressionSet format*

Description

Get datasets (a 'series' object in Gene Expression Omnibus terminology) for a given series GSE identifier for every platform it contains

Usage

```
getDatasets(gse, norm="ORIGINAL", genes=FALSE);
```

Arguments

`gse`	Valid series GSE identifier.
`norm`	Required preprocessing method. Currently available options are ORIGINAL and FRMA. The ORIGINAL normalization is how the original authors submitted it to Gene Expression Omnibus (GEO). This data is always available and is therefore the default option. The availability of FRMA normalization depends on the availability of the CEL files by the original authors.
`genes`	By default a gene expression matrix containing probes is returned. Precomputed conversions using genes instead of probes are also available if this argument is set to TRUE. This conversion was made using the `nsFilter` function from Bioconductors `genefilter` package.

Value

A list with a Bioconductors ExpressionSet for every platform

Examples

```
## Not run:
  getPlatforms("GSE781");
  ## [1] "GPL96" "GPL97"
  esets = getDatasets("GSE781");
  sapply(esets, annotation)
  ## [1] "hgu133a" "hgu133b"

## End(Not run)
```

`getPlatforms` *Get list of platforms for given dataset*

Description

Get list of platforms for given dataset

Usage

```
getPlatforms(gse);
```

Arguments

`gse`	Valid series GSE identifier.

Value

List of valid platform GPL identifiers.

Examples

```
## Not run:
getPlatforms("GSE4635");
## [1] "GPL96"

## End(Not run)
```

Index