

ArrayExpress

April 19, 2009

ArrayExpress

R objects from ArrayExpress database

Description

ArrayExpress produces an AffyBatch, an ExpressionSet or an NChannelSet from a raw data set from the ArrayExpress database. ArrayExpress needs an Internet connection.

Usage

```
ArrayExpress(input, path = ".", save = TRUE, rawcol = NULL)
```

Arguments

input	an ArrayExpress identifier. The package currently covers 96% of the raw data sets of the database and is not currently dealing with the processed data.
path	the name of the directory in which the files downloaded on the ArrayExpress repository will be extracted. The default is the current directory.
save	if TRUE, the files downloaded from the database will not be deleted from path after executing the function.
rawcol	by default, for the raw data, the columns are automatically selected according to the scanner type. If the scanner is unknown or if the user wants to use different columns than the default, the argument 'rawcol' can be set. For two colour arrays it must be a list with the fields 'R', 'G', 'Rb' and 'Gb' giving the column names to be used for red and green foreground and background. For one colour arrays, it must be a character string with the column name to be used. These column names must correspond to existing column names of the expression files.

Value

The output is an object of class [AffyBatch](#) or [ExpressionSet](#) or [NChannelSet](#) with the raw expression values in the assayData of the object, the information contained in the sdrf file in the phenoData, the adf file in the featureData, and the idf file content in the experimentData.

If several array designs are used in the data set, the output is a list with an object for each array design.

Author(s)

Audrey Kauffmann Maintainer: <audrey@ebi.ac.uk>

See Also

[queryAE](#), [getAE](#), [magetab2bioc](#)

Examples

```
ETABM25.affybatch = ArrayExpress(input = "E-TABM-25")
print(ETABM25.affybatch)
sampleNames(ETABM25.affybatch)
colnames(pData(ETABM25.affybatch))
```

extract.zip

Unzip archives in a specified directory

Description

extract.zip extracts the files from a .zip archive in a specific directory.

Usage

```
extract.zip(file, unzip = getOption("unzip"))
```

Arguments

file	A file name.
unzip	character string: the method to be used, an empty string indicates "internal".

Value

Success is indicated by returning the directory in which the files have been extracted. If it fails, it returns an empty character string.

Author(s)

Audrey Kauffmann Maintainer: <audrey@ebi.ac.uk>

getAE	<i>Download MAGE-TAB from ArrayExpress in a specified directory</i>
-------	---

Description

getAE downloads and extracts the MAGE-TAB files from an ArrayExpress data set.

Usage

```
getAE(input, path = ".", save = TRUE, type = "full")
```

Arguments

input	is an ArrayExpress identifier.
path	is the name of the directory in which the files downloaded on the ArrayExpress repository will be extracted.
save	if TRUE, the files downloaded from the database will not be deleted from path after executing the function.
type	can be 'raw' to download and extract only the raw data, 'processed' to download and extract only the processed data or 'full' to have both raw and processed data.

Value

A list with the names of the files that have been downloaded and extracted.

Author(s)

Audrey Kauffmann Maintainer: <audrey@ebi.ac.uk>

See Also

[ArrayExpress](#), [magetab2bioc](#), [queryAE](#)

Examples

```
TABM25 = getAE("E-TABM-25", type = "raw")
```

magetab2bioc	<i>Convert TAB-MAGE data into a Bioconductor object</i>
--------------	---

Description

magetab2bioc converts local MAGE-TAB files into a AffyBatch, an ExpressionSet or a NChannelSet.

Usage

```
magetab2bioc(rawfiles, sdrf, idf, adf, path = ".", rawcol = NULL, save = TRUE)
```

Arguments

rawfiles	all the expression files to use to create the object. The content of the raw.zip MAGE-TAB file.
sdrf	the sdrf file from MAGE-TAB.
idf	the idf file from MAGE-TAB.
adf	the adf file from MAGE-TAB.
path	is the name of the directory in which the files downloaded on the ArrayExpress repository will be extracted.
rawcol	by default, the columns are automatically selected according to the scanner type. If the scanner is unknown or if the user wants to use different columns than the default, the argument 'rawcol' can be set. For two colour arrays it must be a list with the fields 'R', 'G', 'Rb' and 'Gb' giving the column names to be used for red and green foreground and background. For one colour arrays, it must be a character string with the column name to be used. These column names must correspond to existing column names of the expression files.
save	if TRUE, the files downloaded from the database will not be deleted from path after executing the function.

Value

An object of class [AffyBatch](#) or [ExpressionSet](#) or [NChannelSet](#) with the raw expression values in the assayData of the object, the information contained in the .sdrf file in the phenoData, the adf file content in the featureData and the idf file content in the experimentData.

If several array designs are used in the data set, the output is a list with an object for each array design.

Author(s)

Audrey Kauffmann Maintainer: <audrey@ebi.ac.uk>

See Also

[queryAE](#), [getAE](#), [ArrayExpress](#)

Examples

```
MEXP1422 = getAE("E-MEXP-1422")
MEXP1422set = magetab2bioc(rawfiles = MEXP1422$rawfiles, sdrf = MEXP1422$sdrf, idf = MEXP1422$idf)
```

queryAE

XML query of the ArrayExpress repository

Description

queryAE queries ArrayExpress with keywords and give a list of ArrayExpress identifiers and their availability as raw and/or processed data, as an output.

Usage

```
queryAE(keywords = NULL, species = NULL)
```

Arguments

keywords the keyword(s) of interest. To use several words, they must be separated by a "+" as shown in the examples.

species the specie(s) of interest.

Value

A dataframe with character strings with all the ArrayExpress dataset identifiers which correspond to the query in the first column. The two following columns are filled with 'Yes' and 'No' according to the existence of raw and processed data.

Author(s)

Audrey Kauffmann Maintainer: <audrey@ebi.ac.uk>

See Also

[ArrayExpress](#), [getAE](#), [magetab2bioc](#)

Examples

```
## To retrieve all the identifiers of cancer data sets
Cancer = queryAE(keywords = "cancer")

## To retrieve all the identifiers of cancer data sets studied in human
CancerHS = queryAE(keywords = "cancer", species = "homo+sapiens")

## To retrieve all the identifiers of breast cancer data sets studied in human
BreastCancerHS = queryAE(keywords = "breast+cancer", species = "homo+sapiens")
```

Index

*Topic **datasets**

- ArrayExpress, 1
- extract.zip, 2
- getAE, 3
- magetab2bioc, 3
- queryAE, 4

AffyBatch, 1, 4

ArrayExpress, 1, 3–5

ExpressionSet, 1, 4

extract.zip, 2

getAE, 2, 3, 4, 5

magetab2bioc, 2, 3, 3, 5

NChannelSet, 1, 4

queryAE, 2, 3, 4, 4