Package ‘ArrayExpress’

May 18, 2026

Title Access the ArrayExpress Collection at EMBL-EBI Biostudies and build Bioconductor data structures: ExpressionSet, AffyBatch, NChannelSet

Version 1.73.0

Author Audrey Kauffmann, Ibrahim Emam, Michael Schubert, Jose Marugan

Maintainer Jose Marugan <`jcmarca@ebi.ac.uk`>

Depends R (>= 2.9.0), Biobase (>= 2.4.0)

Imports oligo, limma, httr, utils, jsonlite, rlang, tools, methods

Description Access the ArrayExpress Collection at EMBL-EBI Biostudies and build Bioconductor data structures: ExpressionSet, AffyBatch, NChannelSet.

License Artistic-2.0

biocViews Microarray, DataImport, OneChannel, TwoChannel

NeedsCompilation no

Suggests affy

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

git_branch devel

git_last_commit 59e2522

git_last_commit_date 2026-04-28

Repository Bioconductor 3.24

Date/Publication 2026-05-18

Contents

`ae2bioc` *Convert MAGE-TAB files from raw data into a Bioconductor object*

Description

`ae2bioc` converts local MAGE-TAB files into a AffyBatch, an ExpressionSet or a NChannelSet.

Usage

```
ae2bioc(mageFiles, dataCols = NULL, drop = TRUE)
```

Arguments

`mageFiles`	A list as given from `getAE` function. Containing the following elements: **rawFiles** all the expression files to use to create the object. The content of the raw.zip MAGE-TAB file. **sdrf** the name of the sdrf file from MAGE-TAB. **idf** the name of the idf file from MAGE-TAB. **adf** the name of the adf file from MAGE-TAB. **path** is the name of the directory containing these files.
`dataCols`	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 'dataCols' 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.
`drop`	if TRUE and only one platform in series, the platform name will be dropped.

Value

An object of class `AffyBatch`, `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 dataset, the output is a list with an object for each array design.

Author(s)

Ibrahim Emam

Maintainer: Jose Marugan <jcmarca@ebi.ac.uk>

See Also

`ArrayExpress`, `queryAE`, `getAE`

Examples

```
# An example can be found in the help of the getAE function.
```

`ArrayExpress` *R objects from ArrayExpress collection*

Description

`ArrayExpress` produces an `AffyBatch`, an `ExpressionSet` or a `NChannelSet` from a raw dataset from the ArrayExpress collection of the Biostudies database. `ArrayExpress` needs an Internet connection.

Usage

```
ArrayExpress(accession, path = tempdir(), save = FALSE, dataCols = NULL, drop = TRUE)
```

Arguments

`accession`	an ArrayExpress experiment identifier.
`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.
`dataCols`	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 'dataCols' 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.
`drop`	if TRUE and only one platform in series, the platform name will be dropped.

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, Ibrahim Emam

Maintainer: Jose Marugan <jcmarca@ebi.ac.uk>

See Also

`queryAE`, `getAE`, `ae2bioc`, `getcolproc`, `procset`

Examples

```
ETABM25.affybatch = ArrayExpress("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, extractpath = dirname(file)[1])
```

Arguments

`file`	A file name.
`extractpath`	A path to define where the files are to be extracted.

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: Jose Marugan <jcmarca@ebi.ac.uk>

`getAE` *Download MAGE-TAB files from the ArrayExpress collection in a specified directory*

Description

`getAE` downloads and extracts the MAGE-TAB files from an ArrayExpress dataset.

Usage

```
getAE(accession, path = getwd(), type = "full", extract = TRUE, sourcedir = path, overwrite = FALSE)
```

Arguments

`accession`	is an ArrayExpress experiment identifier.
`path`	is the name of the directory in which the files downloaded on the ArrayExpress repository will be extracted.
`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.
`extract`	if FALSE, the files are not extracted from the zip archive.
`sourcedir`	when local = TRUE, files will be read from this directory.
`overwrite`	if TRUE, overwrite files if they already exist in path, default FALSE.

Value

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

Author(s)

Ibrahim Emam, Audrey Kauffmann

Maintainer: Jose Marugan <jcmarca@ebi.ac.uk>

See Also

`ArrayExpress`, `ae2bioc`, `getcolproc`, `procset`

Examples

```
mexp21 = getAE("E-MEXP-21", type = "full")

## Build a an ExpressionSet from the raw data
MEXP21raw = ae2bioc(mageFiles = mexp21)

## Build a an ExpressionSet from the processed data
cnames = getcolproc(mexp21)
MEXP21proc = procset(mexp21, cnames[2])
```

`getcolproc` *Return the possible column names from processed MAGE-TAB files*

Description

`getcolproc` extracts the column names from processed MAGE-TAB and return them. The output is needed to call the function `procset`.

Usage

```
getcolproc(files)
```

Arguments

`files`	A list as given from `getAE` function. Containing the following elements: **procfile** procfile is the name of the processed MAGE-TAB file to be read. **path** is the name of the directory where to find this file.

Author(s)

Audrey Kauffmann

Maintainer: Jose Marugan <jcmarca@ebi.ac.uk>

See Also

`ArrayExpress`, `queryAE`, `getAE`, `procset`

`getcolraw` *Return the possible column names from raw MAGE-TAB files*

Description

`getcolraw` extracts the column names from raw MAGE-TAB and return them. The output can be use to set the argument 'rawcol' of the function `magetab2bioc`.

Usage

```
getcolraw(rawfiles)
```

Arguments

`rawfiles`	rawfiles are the name of the raw MAGE-TAB files to be read.

Author(s)

Audrey Kauffmann

Maintainer: Jose Marugan <jcmarca@ebi.ac.uk>

See Also

`ArrayExpress`, `queryAE`, `getAE`

`procset` *Convert processed MAGE-TAB files into a Bioconductor object*

Description

`procset` converts local MAGE-TAB files into an `ExpressionSet`.

Usage

```
procset(files, procol)
```

Arguments

`files`	is the list with the names of the processed, the sdrf, the adf and the idf files and the path of the data as given by `getAE`.
`procol`	the name of the column to be extracted from the file. Obtained using `getcolproc`.

Author(s)

Ibrahim Emam, Audrey Kauffmann

Maintainer: Jose Marugan <jcmarca@ebi.ac.uk>

See Also

`queryAE`, `getAE`, `getcolproc`

Examples

```
# An example can be found in the help of the getAE function.
```

`queryAE` *Query of the ArrayExpress collection in Biostudies*

Description

`queryAE` queries the ArrayExpress collection with keywords and give a dataframe with ArrayExpress identifiers and related information, 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 all the ArrayExpress dataset identifiers which correspond to the query in the first column. The following columns contain information about these datasets, such as the number of files, the release date on the database, the title, the author and content.`

Author(s)

Ibrahim Emam, Audrey Kauffmann

Maintainer: Jose Marugan <jcmarca@ebi.ac.uk>

See Also

`ArrayExpress`, `getAE`

Examples

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

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

Index