Data for DExMA package

Juan Antonio Villatoro-García [1,2] **and Pedro Carmona-Sáez** [*] [1,2]

[1]Department of Statistics and Operational Research. University of Granada
[2]Bioinformatics Unit. GENYO, Centre for Genomics and Oncological Research

[*]pedro.carmona@genyo.es

April 17, 2025

packageVersionDExMAdata 1.16.0

Contents

1   Package contents

```
> library(DExMAdata)
> data(IDsDExMA)
> data(SynonymsDExMA)
> data(avaliableIDs)
> data(avaliableOrganism)
> data(DExMAExampleData)
```

Firstly, **DExMAExampleData** contains the objects required to perform the **DExMA** package examples:

- **listMatrixEX**: a list of four expression matrices. The first two matrices contain 200 genes annotated in entrez and the other two contains 175 genes annotated in Offical Gene Symbol.

```
> class(listMatrixEX)

[1] "list"

> head(listMatrixEX$Study1)

          Sample1  Sample2  Sample3   Sample4
100859927 5.439524 6.253319 2.926444 4.4304023
8086      5.769823 5.971453 1.831349 4.0466288
8212      7.558708 5.957130 2.365252 3.4352889
65985     6.070508 7.368602 2.971158 3.7151784
729522    6.129288 5.774229 3.670696 3.9171749
13        7.715065 7.516471 1.349453 0.3390772
```

- **listPhenodatas**: a list of four phenodatas corresponding to the four expression matrices of the listMatrixEX object

```
> class(listPhenodatas)

[1] "list"

> head(listPhenodatas$Study1)

        condition gender     organism race
Sample1  Diseased Female Homo Sapiens   AA
Sample2  Diseased Female Homo Sapiens   AA
Sample3   Healthy Female Homo Sapiens   AA
Sample4   Healthy Female Homo Sapiens    H
```

- **listExpressionSets**: a list of four ExpressionSets objects. It contains the same information as *listMatrixEX* and *listPhenodatas* objects.

```
> class(listExpressionSets)

[1] "list"

> listExpressionSets$Study1

ExpressionSet (storageMode: lockedEnvironment)
assayData: 200 features, 4 samples
  element names: exprs
```

```
protocolData: none
phenoData
  rowNames: Sample1 Sample2 Sample3 Sample4
  varLabels: condition gender organism race
  varMetadata: labelDescription
featureData: none
experimentData: use 'experimentData(object)'
Annotation:
```

- **ExpressionSetStudy5**: an ExpressionSet object similar to the ExpressionSets objects of *listExpression*

```
> class(ExpressionSetStudy5)

[1] "ExpressionSet"
attr(,"package")
[1] "Biobase"

> ExpressionSetStudy5

ExpressionSet (storageMode: lockedEnvironment)
assayData: 200 features, 6 samples
  element names: exprs
protocolData: none
phenoData
  rowNames: newSample1 newSample2 ... newSample6 (6 total)
  varLabels: condition gender organism race
  varMetadata: labelDescription
featureData: none
experimentData: use 'experimentData(object)'
Annotation:
```

- **maObjectDif**: the meta-analysis object (*ObjectMA*) created from the listMatrixEX adn phenodatas objects information. An *ObjectMA* is the object type used in the DExMA package. This type of object is better explained in the DExMA package.

```
> str(maObjectDif)

List of 4
 $ Study1:List of 2
  ..$ mExpres  : num [1:200, 1:4] 5.44 5.77 7.56 6.07 6.13 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:200] "100859927" "8086" "8212" "65985" ...
  .. .. ..$ : chr [1:4] "Sample1" "Sample2" "Sample3" "Sample4"
  ..$ condition: num [1:4] 1 1 0 0
 $ Study2:List of 2
  ..$ mExpres  : num [1:200, 1:6] 4.37 5.94 5.29 5.69 5.73 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:200] "100859927" "8086" "8212" "65985" ...
  .. .. ..$ : chr [1:6] "Sample5" "Sample6" "Sample7" "Sample8" ...
  ..$ condition: num [1:6] 1 1 1 0 0 0
 $ Study3:List of 2
  ..$ mExpres  : num [1:175, 1:4] 4.5 7.24 6.04 4.96 6.15 ...
  .. ..- attr(*, "dimnames")=List of 2
```

```
.. .. ..$ : chr [1:175] "AAA4" "AAAS" "AABT" "AACS" ...
.. .. ..$ : chr [1:4] "Sample11" "Sample12" "Sample13" "Sample14"
..$ condition: num [1:4] 1 1 0 0
$ Study4:List of 2
..$ mExpres  : num [1:175, 1:6] 5.37 5.04 6.84 6.94 6 ...
.. ..- attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:175] "AAA4" "ADRACALIN" "AABT" "ACSF1" ...
.. .. ..$ : chr [1:6] "Sample15" "Sample16" "Sample17" "Sample18" ...
..$ condition: num [1:6] 1 1 1 0 0 0
```

- **maObjetc**: an *ObjectMA* after setting all the studies of *maObjectDif* in Official Gene Symbol annotation.

```
> str(maObject)

List of 4
 $ Study1:List of 2
  ..$ mExpres  : num [1:144, 1:4] 5.44 5.77 7.56 6.07 6.13 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:144] "AAA4" "AAAS" "AABT" "AACS" ...
  .. .. ..$ : chr [1:4] "Sample1" "Sample2" "Sample3" "Sample4"
  ..$ condition: num [1:4] 1 1 0 0
 $ Study2:List of 2
  ..$ mExpres  : num [1:144, 1:6] 4.37 5.94 5.29 5.69 5.73 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:144] "AAA4" "AAAS" "AABT" "AACS" ...
  .. .. ..$ : chr [1:6] "Sample5" "Sample6" "Sample7" "Sample8" ...
  ..$ condition: num [1:6] 1 1 1 0 0 0
 $ Study3:List of 2
  ..$ mExpres  : num [1:175, 1:4] 6.37 6.28 4.59 4.6 4.59 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:175] "AARS1" "AATF" "ABCC2" "ABCD1P4" ...
  .. .. ..$ : chr [1:4] "Sample11" "Sample12" "Sample13" "Sample14"
  ..$ condition: num [1:4] 1 1 0 0
 $ Study4:List of 2
  ..$ mExpres  : num [1:175, 1:6] 5.79 5.59 4.61 4.6 4.63 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:175] "AARS1" "AATF" "ABCC2" "ABCD1P4" ...
  .. .. ..$ : chr [1:6] "Sample15" "Sample16" "Sample17" "Sample18" ...
  ..$ condition: num [1:6] 1 1 1 0 0 0
```

On the other hand, **IDsDExMA** and **SynonymsDExMA** are the neccessary objects to be able to apply the *allSameID()* function of the package *DExMA*.

IDsDExMA is a dataframe that contains the equivalences between the different types of IDs. It also contains a column with the organism to which the annotation refers.

```
> class(IDsDExMA)

[1] "data.frame"

> length(IDsDExMA)

[1] 4
```

```
> names(IDsDExMA)

[1] "GeneSymbol" "Entrez"     "Ensembl"     "Organism"

> head(IDsDExMA$Entrez)

[1] "53288" "27777" "27778" "71661" "76253" "78297"

> head(IDsDExMA$Genesymbol)

NULL

> class(SynonymsDExMA)

[1] "data.frame"

> head(SynonymsDExMA)

       Name GeneSymbol    Organism
1       A1m        Pzp Mus musculus
2  AI893533        Pzp Mus musculus
4       MAM        Pzp Mus musculus
5       Pzp        Pzp Mus musculus
9     Nat-2      Aanat Mus musculus
10   AA-NAT      Aanat Mus musculus
```

SynonymsDExMA is a data.frame of 3 columns that contains other possible names that a gene can have in a organism and its equivalent annotation in Official Gene Symbol.

```
> class(SynonymsDExMA)

[1] "data.frame"

> head(SynonymsDExMA)

       Name GeneSymbol    Organism
1       A1m        Pzp Mus musculus
2  AI893533        Pzp Mus musculus
4       MAM        Pzp Mus musculus
5       Pzp        Pzp Mus musculus
9     Nat-2      Aanat Mus musculus
10   AA-NAT      Aanat Mus musculus
```

avaliableIDs is a character vector that contains the different IDs that are avaliable to use in *allSameID* function. It is recommended to look this object before making use of *allSameID* function.

```
> avaliableIDs

[1] "Entrez"     "Ensembl"     "GeneSymbol"
```

avaliableOrganism is a character vector that contains the different organism that are avaliable to use in *allSameID* function. Like **avaliableIDs** object, it is recommended to look this object before making use of *allSameID* function.

```
> avaliableOrganism

 [1] "Bos taurus"               "Caenorhabditis elegans"
 [3] "Canis familiaris"         "Danio rerio"
```

```
 [5] "Drosophila melanogaster"  "Gallus gallus"
 [7] "Homo sapiens"             "Mus musculus"
 [9] "Rattus norvegicus"        "Arabidopsis thaliana"
[11] "Saccharomyces cerevisiae" "Escherichia coli"
```

2 Sources

listMatrixEX, **lisPhenodatas**, **listExpressionSets** and **ExpressionSetStudy5** example objects are synthetic.

The **maObjectDif** example object have been created after applying *createObjectMA()* function from **DExMA** package to *listMatrixEX* and *listPhenodatas* objects.

The **maObject** example object have been obtained after applying *allSameID()* function from **DExMA** package to *maObjectDif*

IDsDExMA and **SynonymsDExMA** objects have been constructed using the information avaliable in NCBI GEO [1] and in the NCBI's gene database [2]

3 Session info

```
R version 4.5.0 RC (2025-04-04 r88126)
Platform: x86_64-pc-linux-gnu
Running under: Ubuntu 24.04.2 LTS

Matrix products: default
BLAS:   /home/biocbuild/bbs-3.21-bioc/R/lib/libRblas.so
LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.12.0  LAPACK version 3.12.0

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
 [3] LC_TIME=en_GB              LC_COLLATE=C
 [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

time zone: America/New_York
tzcode source: system (glibc)

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] DExMAdata_1.16.0

loaded via a namespace (and not attached):
 [1] digest_0.6.37      fastmap_1.2.0       xfun_0.52
 [4] knitr_1.50         BiocGenerics_0.54.0 Biobase_2.68.0
```

```
 [7] htmltools_0.5.8.1   rmarkdown_2.29      generics_0.1.3
[10] cli_3.6.4           compiler_4.5.0      tools_4.5.0
[13] evaluate_1.0.3      yaml_2.3.10         BiocManager_1.30.25
[16] rlang_1.1.6         BiocStyle_2.36.0
```

References

[1] Barret T., Wilhite S., Ledoux P. and et al. *NCBI GEO: archive for functional genomics data sets–update Nucleic Acids Research*, 991-995, 2013
https://doi.org/10.1093/nar/gks1193

[2] *Database resources of the National Center for Biotechnology information Nucleic Acids Research*, volume 47, Pages D23–D28, 2019 https://doi.org/10.1093/nar/gky1069