Differential Expression Meta-Analysis with DExMA package

Juan Antonio Villatoro-García¹,² and Pedro Carmona-Sáez*

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

*pedro.carmona@genyo.es

December 1, 2023

Abstract

DExMA (Differential Expression Meta-Analysis) performs all the necessary steps of differential expression meta-analysis considering the possible existence of missing genes. It controls the existence of unmeasured genes by two approaches: considering the genes that are contained in a proportion of datasets (set by the users) or imputing the expression values of the unmeasured genes using the Knn method in the space of samples. In addition, it allows to apply quality controls, download GEO datasets and show a graphical representation of the results.

Contents

1 Introduction .......................................................... 3
2 Previous steps: Meta-analysis object ................................. 3
   2.1 Meta-analysis object creation (objectMA) ......................... 4
   2.2 Adding a new dataset to the meta-analysis object ............... 7
3 Performing Meta-analysis .............................................. 9
   3.1 Gene annotation and quality controls .............................. 9
      3.1.1 Setting all the datasets in the same annotation .......... 9
      3.1.2 Logarithm transformation .................................. 11
      3.1.3 Heterogeneity study ....................................... 12
   3.2 Imputing missing genes expression values (optional) .......... 13
   3.3 Performing meta-analysis: metaAnalysisDE() .................. 14
      3.3.1 Effects size combination results ............................ 16
      3.3.2 P-value combination results ................................ 17
   3.4 Visualization of the results: heatmap ............................ 18
4 Additional information ...................................................... 23
  4.1 GEO microarray data download .................................... 23
  4.2 Using RNA-Seq data .................................................. 24
  4.3 Removing Batch Effects .............................................. 24
  4.4 Calculating Effects size ............................................... 26
  4.5 Calculating Individual P-values ................................... 27

5 Session info ................................................................. 27
# Introduction

**DExMA** is a package designed to perform gene expression meta-analysis. Gene expression meta-analysis comprises a set of methods that combine the results of several differential expression studies into a single common result [1]. Furthermore, this package has the advantage that it allows to take into account those genes that are not measured in a study. The first approach to control these missing genes, **DExMA** offers two options: apply the k-nearest neighbors method (knn method) in the space of samples to impute the expression values of these missing genes [2]. The second option is to take into account those genes that are contained in at least a certain proportion (set by the user) of datasets. The use of only the genes common to all the datasets could lead to the loss of some genes that are not measured in a single study, which would lead to the loss of information. Due to this fact, **DExMA** package is very useful to work with Microarray data, because this type of data is what usually produces the existence of uncommon genes between datasets with different annotations. However, previously normalized RNA-Seq data can also be used, as well as, both Microarray data and normalized RNA-Seq data at the same time. **DExMA** package has implemented methods from the three main types of gene expression meta-analysis [1] meta-analysis based on effects sizes combination and meta-analysis based on p-values combination. Once one of the methods has been applied, **DExMA** package provides a specific and adapted results table. Moreover, this package contains some functions that allows the user to carry out a previous quality control in order to results obtained are more reliable. Finally, this package provides some additional function that, for example, help the user to download public Microarray data from NCBI GEO public database [3] or to visualize the significant genes in a heatmap.

This document gives a tutorial-style introduction to all the steps that must be carry out in order to properly perform gene expression meta-analysis by making use of **DExMA** package.

## Previous steps: Meta-analysis object

**DExMA** uses a specific object as input, which is a list of nested lists where each nested list corresponds to a study. This object can be created directly by the users or they can use `createObjectMA()` function to create it.

For the examples that are going to be shown, synthetic data will be used. We load the sample data into our R session.

```r
> library(DExMA)
> data("DExMAExampleData")
```

- `listMatrixEX`: a list of four expression arrays
- `listPhenodatas`: a list of the four phenodata corresponding to the four expression arrays
- `listExpression`: a list of four ExpressionSets object. It contains the same information as `listMatrixEX` and `listPhenodatas`
- `ExpressionSetStudy5`: an ExpressionSet object similar to the ExpressionSets objects of `listExpression`
- `maObjectDif`: the meta-analysis object created from the `listMatrixEX` and `listPhenodatas` objects.
DExMA package

- maObject: the meta-analysis object after setting all the studies in Official Gene Symbol annotation

2.1 Meta-analysis object creation (objectMA)

As previously stated, the meta-analysis input in DExMA is a list of nested lists. Each nested list contains two elements:

- A gene expression matrix with genes in rows and samples in columns
- A vector of 0 and 1 indicating the group of each sample. 0 represents reference group (usually controls) and 1 represents experimental group (usually cases).

This object can be created directly by the user or we can make use of createObjectMA() function, which creates the *objectMA* after indicating how the reference and experimental groups are identified.

createObjectMA() function allows to create the object needed to perform meta-analysis. In this case, it is necessary to indicate as input of the function the variables that contain the experimental and reference groups:

- listEX: a list of dataframes or matrix (genes in rows and samples in columns). A list of ExpressionSets can be used too:

```r
> # List of expression matrices
> data("DExMAExampleData")
> ls(listMatrixEX)
[1] "Study1" "Study2" "Study3" "Study4"
> 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
```

```r
> # List of ExpressionSets
> ls(listExpressionSets)
[1] "Study1" "Study2" "Study3" "Study4"
> 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:

- **listPheno**: a list of phenoDatas (samples in rows and covariables in columns). If the object listEX is a list of ExpressionSets this element can be null.

```r
> data("DExMAExampleData")
> #Example of a phenodata object
> ls(listPhenodatas)
[1] "Study1" "Study2" "Study3" "Study4"
> 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
```

- **namePheno**: a list or vector of the different column names or column positions from the pheno used for performing the comparison between groups. Each element of namePheno correspond to its equivalent element in the listPheno. (default a vector of 1, all the first columns of each elements of listPheno are selected)

- **expGroups**: a list or vector of the group names or positions from namePheno variable used as experimental group (cases) to perform the comparison (default a vector of 1, all the first groups are selected).

- **refGroups**: a list or vector of the group names or positions from namePheno variable used as reference group (controls) to perform the comparison (default a vector of 2, all the second groups are selected).

It is important to note that if any element does not belong to the experimental or the reference group, that sample is not taken into account in the creation of meta-analysis object.

Here, we have included an example to show how exactly the function is used:

Since this function can be a bit complicated if there are many datasets, we recommend creating a vector to keep the column names of the phenodatas that contains the variable that identifies the groups to compare (namePheno argument). Moreover, we should create two others lists to indicate how to identify experimental (cases) and reference (controls) groups in these variables (expGroups and refGroups arguments).

If we look at the example phenodatas list we have the following four objects:

```r
> 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
```

In the "Study1" phenoData, the groups variable is "condition". Experimental group is named as "Diseased" and reference group as "Healthy".
In the "Study2" phenoData, the groups variable is "condition". Experimental group is named as "Diseased" or "ill" and reference group as "Healthy" or "control".

In the "Study3" phenoData, the groups variable is "state". Experimental group is named as "Diseased" and reference group as "Healthy".

In this phenoData, the groups variable is "state". Experimental group is named as "ill" and reference group as "control".

We all this information we can create the vector for namePheno argument and the two list for expGroups and refGroups:

```r
> phenoGroups = c("condition", "condition", "state", "state")
> phenoCases = list(Study1 = "Diseased", Study2 = c("Diseased", "ill"),
+                     Study3 = "Diseased", Study4 = "ill")
> phenoControls = list(Study1 = "Healthy", Study2 = c("Healthy", "control"),
+                      Study3 = "Healthy", Study4 = "control")
```

Then, we can apply more easily createObjectMA() function:

```r
> newObjectMA <- createObjectMA(listEX=listMatrixEX,
+                               listPheno = listPhenodatas,
+                               namePheno=phenoGroups,
+                               expGroups=phenoCases,
+                               refGroups = phenoControls)
> #Study 1
```
DExMA package

> head(newObjectMA[[1]][[1]])
          Sample1  Sample2  Sample3  Sample4
8086     5.769823  5.971453  2.971158  3.715178
8212     7.558708  5.957130  3.435289
65985    6.070508  7.368602  2.971158  3.715178
729522   6.129288  5.774229  3.917174
13       7.715065  7.516471  1.349453  0.339077

> newObjectMA[[1]][[2]]
[1] 1 1 0 0

> #Study 2
> head(newObjectMA[[2]][[1]])
          Sample5  Sample6  Sample7  Sample8  Sample9  Sample10
8086      5.937004  5.914434  3.234439  3.228559  4.521235
8212      5.294553  4.092231  3.683300  2.750491  5.369983
65985      5.685822  5.900363  3.153294  3.166798  3.162232
729522     5.733054  5.539666  2.750492  2.659061  2.452524
13        6.153159  5.435909  4.860632  3.751608  3.121853

> newObjectMA[[2]][[2]]
[1] 1 1 1 0 0 0

The result obtained is the proper object to perform meta-analysis (objectMA).

2.2 Adding a new dataset to the meta-analysis object

It may happen that once the meta-analysis object is created we want to add a new dataset before doing the meta-analysis. DExMA provides the `elementObjectMA()` function, which allows the creation of an element of the meta-analysis object. This function contains the following arguments:

- `expressionMatrix`: a dataframe or matrix that containing genes in rows and samples in columns. An ExpressionSet object can be used too.
- `pheno`: a data frame or a matrix containing samples in rows and covariates in columns. If NULL (default), pheno is extracted from the ExpressionSet object
- `groupPheno`: the column name or position from pheno where experimental group (cases) and reference group (control) are identified
- `expGroups`: a vector of the names or positions from groupPheno variable used as experimental group (cases). By default the first group (character) is taken
- `refGroups`: a vector of the names or positions from groupPheno variable used as reference group (control). By default the second group (character) is taken.

As with the `createObjectMA()` function, if any element does not belong to the experimental or the reference group, that sample is not included in the creation of the object.

Here we provided an example of the use of this function, in which we create an element of the meta-analysis object from the information of the "Study 2" of the `listExpressionSets` object:
We had to load the Biobase package in order to extract the information in the ExpressionSet. In the phenoData we can see that the groups variable is "condition". In addition, Experimental group is name as "Diseased" or "ill" and reference group as "Healthy" or "control".

```r
> newElem <- elementObjectMA(expressionMatrix = ExpressionSetStudy5,
+  groupPheno = "condition",
+  expGroup = c("Diseased", "ill"),
+  refGroup = c("Healthy", "control"))
> head(newElem[[1]])
newSample1  newSample2  newSample3  newSample4  newSample5  newSample6
 10085927     5.867690     8.148252     7.286897     4.320073     4.188687     4.831881
  8086       7.437004     7.414434     8.624650     4.734439     4.728559     6.021235
 65985       7.185822     7.400363     7.039580     3.653294     4.666798     4.662232
 729522      7.233054     7.053966     8.092146     4.775042     4.159061     3.952524
     13      7.653159     6.935909     8.491360     6.360632     5.251608     4.621853
```

As we can see, we obtain a list that has the same structure as the elements of the meta-analysis object (objectMA). This new element can be added to a objectMA that has been created previously:

```r
> newObjectMA2 <- newObjectMA
> newObjectMA2[[5]] <- newElem
```
Moreover, an advantage of this function is that it can be used to create one by one all the elements and finally join all of them to create the meta-analysis object.

3 Performing Meta-analysis

**DExMA** package contains the main gene expression meta-analysis methods:

- Meta-analysis based on effect size combination: Fixed Effects Model (FEM) and Random Effects Model (REM).
- Meta-analysis based on P-value combination: Fisher’s method, Stouffer’s method, Wilkonson’s method (maxP), Tippett’s method (minP) and Aggregated Cauchy Association Test method (ACAT).

These methods can be applied directly, but it is advisable to apply some previous **DExMA** function to ensure the results are accurate.

3.1 Gene annotation and quality controls

Before performing the meta-analysis, all the genes must be in the same annotation and a quality control should be done in order to obtain reliable results [1]. **DExMA** provides some useful functions to help the user to do it.

3.1.1 Setting all the datasets in the same annotation

All genes must be in the same annotation in order to perform the meta-analysis successfully and avoid incorrect interpretations of the results. In cases where all datasets do not have the same gene ID, **DExMA** contains a function called `allSameID()` that allows to annotated all the datasets with the same gene ID. Specifically, the function allows to annotate those datasets that use the Official Gene Symbol, Entrez or Ensembl. The inputs of this function are:

- `objectMA`: the meta-analysis object of **DExMA** package. The result obtained by `createObjectMA()` function should be used.
- `finalID`: a character that indicates the final gene ID that all the studies will have.
- `organism`: a character that indicates the organism that the datasets belong to. Use `availableOrganism` function to see the organism admitted.
Here we include an example of how to use this function. In this example we have used "newObjectMA" that has been created before. The first two expression arrays are annotated in "entrez", the third expression matrix in "Official Gene Symbol" and the last one in "Official Gene Symbol" with some synonyms:

```r
> rownames(newObjectMA$Study1$mExpres)[1:20]
[1] "100859927" "8086" "8212" "65985" "729522" "13"
[7] "344752" "126767" "343066" "51166" "79719" "22848"
[13] "14" "15" "16" "57505" "80755" "132949"
[19] "60496" "10157"
> rownames(newObjectMA$Study2$mExpres)[1:20]
[1] "100859927" "8086" "8212" "65985" "729522" "13"
[7] "344752" "126767" "343066" "51166" "79719" "22848"
[13] "14" "15" "16" "57505" "80755" "132949"
[19] "60496" "10157"
> rownames(newObjectMA$Study3$mExpres)[1:20]
[1] "AAA4" "AAAS" "AABT" "AACS" "AACSP1" "AADAC"
[7] "AADACL2" "AADACL3" "AADACL4" "AADAT" "AAGAB" "AAK1"
[13] "AAMP" "AANAT" "AARS1" "AARS2" "AARSD1" "AASDH"
[19] "AASDHPPPT" "AASS"
> rownames(newObjectMA$Study4$mExpres)[1:20]
[1] "AAA4" "ADRACALIN" "AABT" "ACSF1" "AACSP1" "AADAC"
[7] "AADACL2" "AADACL3" "AADACL4" "AADAT" "PPKP1" "AAK1"
[13] "AAMP" "AANAT" "AARS1" "AARS2" "AARSD1" "AASDH"
[19] "AASDHPPPT" "AASS"
```

We can use `avaliableIDs` and `avaliableOrganism` in order to know how to write in the function the `finalID` and the `organism` arguments:

```r
> head(avaliableIDs)
[1] "Entrez" "Ensembl" "GeneSymbol"
> 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"
```

Then, we are ready to run the `allSameID()` function. We are going to annotated all the datasets in Official Gene Symbol (`finalID="GeneSymbol"`):

```r
> newObjectMA <- allSameID(newObjectMA, finalID="GeneSymbol", organism = "Homo sapiens")
```

<table>
<thead>
<tr>
<th></th>
<th>0%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25%</td>
</tr>
</tbody>
</table>
As it can be seen, all the studies are now annotated in Official Gene Symbol.

### 3.1.2 Logarithm transformation

To avoid problems with the returned fold-change by the meta-analysis, $\log_2$ should be applied to the gene expression values. We can make use of `dataLog()` function to check if each dataset expression values have the $\log_2$ applied already. If not, the function will make the transformation:

```r
> newObjectMA <- dataLog(newObjectMA)
> head(newObjectMA$[[1]][[1]])
Sample1  Sample2  Sample3  Sample4
AAA4     5.439524  6.253319  2.926444  4.4304023
AAAS     5.769823  5.971453  1.831349  4.0466288
AABT     7.558708  5.957130  2.365252  3.4352889
AACS     6.070508  7.368602  2.971158  3.7151784
AACSP1   6.129288  5.774229  3.670696  3.9171749
AADAC    7.715065  7.516471  1.349453  0.3390772
```

As it can be seen, all the studies are now annotated in Official Gene Symbol.
### 3.1.3 Heterogeneity study

Some heterogeneity between studies may lead to some methods, such as the Fixed Effects Model, not providing reliable results. Therefore, it is advisable to carry out a study of heterogeneity to correctly choose the meta-analysis method [1]. The `heterogeneityTest()` function shows two ways of measuring heterogeneity.

On the one hand, it returns a QQ-plot of the Cochran’s test [4]. In this plot, if most of the values are close to the central line, that is, most of the Cochran’s test values are close to the expected distribution (chi-squared distribution), it can be said that there is homogeneity. In the case that these values deviate greatly from the expected distribution, it must be assumed that there is heterogeneity.

On the other hand, $I^2$ measures the percentage of variation across studies due to heterogeneity [5]. In the case of gene expression data, an $I^2$ for each gene across datasets would have to be calculated. As in the case of many genes, it can be difficult to observe all the $I^2$ values obtained, the `heterogeneityTest()` function returns the quantiles of the different $I^2$ values calculated. $I^2$ values equal to 0 indicate homogeneity and values less than 0.25 are usually categorized as low heterogeneity [5]. Therefore, to assume homogeneity in the gene expression meta-analysis, almost all $I^2$ values must be 0 or at least less than 0.25.

In the example shown below, it is observed that in the QQ-plot of the Cochran’s test, Q-values deviate considerably from the expected distribution and approximately 10% of the $I^2$ values are greater than 0.25, therefore homogeneity could not be assumed.

```r
> heterogeneityTest(newObjectMA)
[1] "I^2 Quantiles"
   0%   25%   50%   75%
0.0000000 0.0000000 0.4484812 0.6865470
```
3.2 Imputing missing genes expression values (optional)

Optionally, the `missGenesImput()` function allows to impute the expression values of the unmeasured genes. To perform this imputation, the function apply the knn method in the space of samples (`sampleKnn`) [2]. Once the function have been applied, it returns the same `objectMA` with all the datasets imputed and some imputation indicators:

```r
> nrow(newObjectMA$Study1[[1]])
[1] 144
> nrow(newObjectMA$Study2[[1]])
[1] 144
> imputation <- missGenesImput(maObject, k =7)
```

```
[1] "Number of values imputed 310 (0.089 %)"
[1] "Number of genes imputed in Study1: 31 of 175 (17.7%)"
[1] "Number of genes imputed in Study2: 31 of 175 (17.7%)"
```
If we observe the number of genes in Study1, once the imputation is performed, we can observe that it goes from having 144 genes to 175.

```r
> nrow(maObject_imput$Study1[[1]])
[1] 175
> nrow(maObject_imput$Study2[[1]])
[1] 175
```

In addition, the function returns some indicators to evaluate the reliability of the imputation:

- `imputValuesSample`: Number of missing values imputed per sample
- `imputPercentageSample`: Percentage of missing values imputed per sample
- `imputValuesGene`: Number of missing values imputed per gene
- `imputPercentageGene`: Percentage of missing values imputed per gene

```r
> head(imputation$imputIndicators$imputValuesSample)
Sample1  Sample2  Sample3  Sample4  Sample5  Sample6
     31       31       31       31       31       31
> head(imputation$imputIndicators$imputPercentageSample)
Sample1  Sample2  Sample3  Sample4  Sample5  Sample6
 17.71%   17.71%   17.71%   17.71%   17.71%   17.71%
> head(imputation$imputIndicators$imputValuesGene)
   AAA4 AAAS  AABT  AACS AACSPI AADAC
     0     0     0     0     0     0
> head(imputation$imputIndicators$imputPercentageGene)
   AAA4 AAAS  AABT  AACS AACSPI AADAC
     0     0     0     0     0     0
```

### 3.3 Performing meta-analysis: `metaAnalysisDE()`

The `metaAnalysisDE()` function allows to perform a meta-analysis in only one step, needing only the meta-analysis object created previously.

This function has as input:

- `objectMA`: The meta-analysis object of DExMA package. The result obtained by `createObjectMA` function should be used.
- `effectS`: A list of three elements. The first element is a dataframe with genes in rows and studies in columns. Each component of the dataframe is the effect of a gene in a study. The second element of the list is also a dataframe with the same structure, but in this case each component of the dataframe represent the
DExMA package

The result of the calculateES function can be used too. The third element of the list is also a dataframe with the same structure, but in this case each component of the dataframe represent the log fold change of a gene in a study. This argument should be only used in the case that objectMA argument is null.

- pvalues: A list of two elements. The first element is a dataframe with genes in rows and studies in columns. Each component of the dataframe is the p-value of a gene in a study. The second element of the list is also a dataframe with the same structure, but in this case each component of the dataframe represent the log fold change of a gene in a study. This argument should be only used in the case that objectMA argument is null.

- weight: A vector of the weights of each dataset. This argument should only be included in case objectMA is null and you want to use "Stouffer" or "ACAT" method.

- typeMethod: a character that indicates the method to be performed:
  - "FEM": Fixed Effects model.
  - "REM": Random Effects model.
  - "Fisher": Fisher’s method (sum of logarithms of p-values)
  - "Stouffer": Stouffer’s method (sum of z-scores)
  - "maxP": Wilkinson’s method (maximum of p-values)
  - "minP": Tippet’s method (minimum of p-values)
  - "ACAT": Aggregated Cauchy Association Test method

- missAllow: a number between 0 and 1 that indicates the maximum proportion of missing values allows in a sample. If the sample has more proportion of missing values, the sample will be eliminated. In the other case, the missing values will be imputed by using the K-NN algorithm included in impute package [6]. In case the objectMA has been previously imputed, this element is not necessary.

- proportionData: a number between 0 and 1 that indicates the minimum proportion of datasets in which a gene must be contained to be included. In case the objectMA has been previously imputed, this element is not necessary.

- Adjusted p-value from which a gene is considered significant. Default 0.05. This value only takes you into account in the results generated in rank combination methods.

In the following example, we have applied a Random Effect model to the DExMA object ("newObjectMA") we have been working with so far. In addition we have allowed a 0.3 proportion of missing values in a sample and a gene must have been contained in at least the 50% of studies.

```r
> resultsMA <- metaAnalysisDE(newObjectMA, typeMethod="REM", + missAllow=0.3, proportionData=0.50)
[1] "Performing Random Effects Model"
```

The output of this function is a dataframe with the results of the meta-analysis where rows are the genes and columns are the different variables provided by the meta-analysis:
### DExMA package

```r
> head(resultsMA)

<table>
<thead>
<tr>
<th></th>
<th>Com.ES</th>
<th>ES.var</th>
<th>Qval</th>
<th>Qpval</th>
<th>tau2</th>
<th>Zval</th>
<th>Pval</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA4</td>
<td>2.176422</td>
<td>0.5769960</td>
<td>7.390096</td>
<td>0.066045064</td>
<td>1.2841381</td>
<td>2.857312</td>
<td>4.272453e-03</td>
</tr>
<tr>
<td>AAAS</td>
<td>2.164285</td>
<td>0.2161334</td>
<td>1.348614</td>
<td>0.71762287</td>
<td>0.0000000</td>
<td>4.655362</td>
<td>3.23418e-06</td>
</tr>
<tr>
<td>AABT</td>
<td>2.502180</td>
<td>0.8198233</td>
<td>7.905694</td>
<td>0.04800146</td>
<td>1.9149389</td>
<td>2.763493</td>
<td>5.718633e-03</td>
</tr>
<tr>
<td>AACS</td>
<td>2.256773</td>
<td>0.4634824</td>
<td>5.424099</td>
<td>0.14324946</td>
<td>0.8093607</td>
<td>3.149067</td>
<td>9.167395e-04</td>
</tr>
<tr>
<td>AACSP1</td>
<td>3.685009</td>
<td>0.4675196</td>
<td>2.393087</td>
<td>0.49403773</td>
<td>0.0000000</td>
<td>5.893789</td>
<td>7.070214e-08</td>
</tr>
<tr>
<td>AADAC</td>
<td>2.641814</td>
<td>0.9830839</td>
<td>9.279712</td>
<td>0.02579411</td>
<td>2.3971453</td>
<td>2.664446</td>
<td>7.711514e-03</td>
</tr>
</tbody>
</table>
```

Moreover, this function can also be applied to the imputed objectMA, but in this case, it is not necessary to indicate the missAllow and proportionData elements:

```r
> resultsMA_imput <- metaAnalysisDE(maObject_imput, typeMethod="REM",)

> head(resultsMA_imput)

<table>
<thead>
<tr>
<th></th>
<th>Com.ES</th>
<th>ES.var</th>
<th>Qval</th>
<th>Qpval</th>
<th>tau2</th>
<th>Zval</th>
<th>Pval</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA4</td>
<td>1.96577e-02</td>
<td>2.734153</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAAS</td>
<td>9.32844e-05</td>
<td>3.163741</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AABT</td>
<td>2.357305e-02</td>
<td>2.746359</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AACS</td>
<td>7.292246e-03</td>
<td>3.009776</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AACSP1</td>
<td>1.237287e-05</td>
<td>3.088231</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AADAC</td>
<td>2.998922e-02</td>
<td>2.897440</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

The variables of the dataframe change from one type of meta-analysis to another. A more detailed explanation of these results will be addressed in the following sections.

#### 3.3.1 Effects size combination results

The "FEM" and "REM" methods provide a dataframe with the variables:

- **Com.ES**: combined effect of the gene.
- **ES.var**: variance of the combined effect of the gene.
- **Qval**: total variance of the gene.
- **Qpval**: p-value for the total variance of the gene.
DExMA package

- tau2: between-study variance of the gene.
- zval: combined effect value for a standard normal. I can be use in order to find out if the gene is overexpressed (positive value) or underexpressed (negative value).
- Pval: P-value of the meta-analysis for the gene.
- FDR: P-value adjusted of the meta-analysis for the gene.
- AveFC: Weighted average of log Fold-Change values for the gene used in order to find out if the gene is overexpressed (positive value) or underexpressed (negative value).
- Prop.dataset: Proportion of the datasets in which the gene is included.

```r
> resultsES <- metaAnalysisDE(newObjectMA, typeMethod="REM", proportionData=0.5)
[1] "Performing Random Effects Model"
> head(resultsES)
   Com.ES ES.var  Qval  Qpval  tau2   Zval  Pval
AAA4 2.170422 0.5769960 7.390096 0.06045064 1.2841381 2.857312 4.272453e-03
AAAS 2.164285 0.2161334 1.348614 0.71762287 0.0000000 4.655362 3.234118e-06
AABT 2.502180 0.8198233 7.905694 0.04800146 1.9149389 2.763493 5.718633e-03
AACS 2.256773 0.4634824 5.424099 0.14324946 0.8093607 3.314906 9.167395e-04
AACSP1 3.685009 0.4675196 2.393007 0.49493773 0.0000000 5.389378 7.070214e-08
AADAC 2.641814 0.9830839 9.279712 0.02579411 2.3971453 2.664446 7.711514e-03
```

3.3.2 P-value combination results

The "Fisher", "Stouffer", "minP", "maxP" and "ACAT" methods provide a dataframe with the following variables:
- Stat: Statistical calculated in the method
- Pval: P-value of the meta-analysis for the gene.
- FDR: P-value adjusted of the meta-analysis for the gene.
- AveFC: Average of log Fold-Change values for the gene used in order to find out if the gene is overexpressed (positive value) or underexpressed (negative value).
- Prop.dataset: Proportion of the datasets in which the gene is included.

Here we present an example making use of "maxP" method:

```r
> resultsPV <- metaAnalysisDE(newObjectMA, typeMethod="maxP", proportionData=0.5)
[1] "Performing MaxP's method"
> head(resultsPV)
   FDR AveFC propDataset
AAA4 1.967577e-02 2.734153 1
AAAS 9.432844e-05 3.163741 1
AABT 2.357305e-02 2.746359 1
AACS 7.292246e-03 3.009776 1
AACSP1 1.237287e-05 3.088231 1
AADAC 2.998922e-02 2.897440 1
```
3.4 Visualization of the results: heatmap

Finally, we can represent in a heatmap the significant genes in order to observe how they are expressed in each of the studies. In `makeHeatmap()` function we have to include both the object that has been used in the meta-analysis, the result of it and the applied method. In addition, this package offers three different scaling approaches (scaling) in order to compare properly the gene expression of the studies in the heatmap:

- "zscor": It calculates a z-score value for each gene, that is, the mean gene expression from each gene is subtracted from each gene expression value and then it is divided by the standard deviation.
- "swr": Scaling relative to reference dataset approach [7].
- "rscale": It uses the rescale function of the `scales` package to scale the gene expression [8].
- "none": no scaling approach is applied.

Moreover, in `regulation` argument, we can choose if we want to represent the overexpressed or underexpressed genes:

- "up": only up-expressed genes are represented.
- "down": only down-expressed genes are represented
- "all": up-expressed and down-expressed genes are represented.

We can choose the number of significant genes (`numSig`) that we want to be shown on the graph and the adjusted p-value from which a gene is considered as significant (`fdrSig`). In addition, the genes that are not presented in one sample are represented in gray.

Here we present an example of the heatmap which have been obtained from the result of applying a random effects model to the object "newObjectMA" and making use of a "zscor" scaling approach.

```r
> makeHeatmap(objectMA=newObjectMA, resMA=resultsMA, scaling = "zscor",
+ regulation = "all", numSig=40,
+ fdrSig = 0.05, logFCSig = 1.5, show_rownames = TRUE)
[1] "scaling using z-score..."
```

Sample1.Study1 Sample2.Study1 Sample3.Study1 Sample4.Study1
AARS2 0.9682999 0.50830923 -1.3357281 -0.1408811
AADACL3 0.1962999 1.27309503 -0.3733564 -1.0960385
AARS1 0.9972642 0.69944418 -0.6523539 -1.0443545
AAAS 0.7106395 0.81561074 -1.3397753 -0.1864749
AACSP1 0.9985590 0.71637534 -0.9554120 -0.7595223
<table>
<thead>
<tr>
<th>Sample</th>
<th>Sample5.Study2</th>
<th>Sample6.Study2</th>
<th>Sample7.Study2</th>
<th>Sample8.Study2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AARS2</td>
<td>1.29003246</td>
<td>-0.1493767</td>
<td>1.1881662</td>
<td>-0.6257615</td>
</tr>
<tr>
<td>AADAACL3</td>
<td>1.18101154</td>
<td>0.9235015</td>
<td>0.5709817</td>
<td>-0.7336264</td>
</tr>
<tr>
<td>AARS1</td>
<td>0.32261014</td>
<td>1.3600903</td>
<td>0.7049359</td>
<td>-1.2920142</td>
</tr>
<tr>
<td>AAAS</td>
<td>0.59178756</td>
<td>0.5776332</td>
<td>1.3366176</td>
<td>-1.1031207</td>
</tr>
<tr>
<td>AACSP1</td>
<td>0.75769816</td>
<td>0.6575860</td>
<td>1.2379418</td>
<td>-0.6163623</td>
</tr>
<tr>
<td>AAGAB</td>
<td>0.66979718</td>
<td>0.5913302</td>
<td>1.3541761</td>
<td>-0.8630518</td>
</tr>
<tr>
<td>AACS</td>
<td>0.87262982</td>
<td>1.0046754</td>
<td>0.7826122</td>
<td>-1.3015653</td>
</tr>
<tr>
<td>AASDHPPPT</td>
<td>1.03567093</td>
<td>0.9364763</td>
<td>0.5862831</td>
<td>-1.4423486</td>
</tr>
<tr>
<td>AADAC</td>
<td>0.75691292</td>
<td>0.2637028</td>
<td>1.3329252</td>
<td>-0.1318817</td>
</tr>
<tr>
<td>AARS1D</td>
<td>0.88971101</td>
<td>0.2844332</td>
<td>1.3444491</td>
<td>-1.1891260</td>
</tr>
<tr>
<td>AABT</td>
<td>0.85037088</td>
<td>-0.2263859</td>
<td>0.7607782</td>
<td>-0.8746852</td>
</tr>
<tr>
<td>AAA1</td>
<td>0.85715809</td>
<td>1.4476910</td>
<td>0.9228507</td>
<td>-0.8864413</td>
</tr>
<tr>
<td>AADAACL2</td>
<td>1.40453880</td>
<td>0.2067201</td>
<td>0.7295036</td>
<td>-0.4744804</td>
</tr>
<tr>
<td>AMP</td>
<td>0.61483783</td>
<td>1.0780356</td>
<td>0.7377432</td>
<td>-1.1325895</td>
</tr>
<tr>
<td>AANAT</td>
<td>0.42906396</td>
<td>0.8208962</td>
<td>1.3123105</td>
<td>-0.7528130</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
<th>Value 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAGAB</td>
<td>1.0769772</td>
<td>0.61814883</td>
<td>-0.9507280</td>
<td>-0.7443980</td>
</tr>
<tr>
<td>AACS</td>
<td>0.5086833</td>
<td>1.14412657</td>
<td>-1.0085114</td>
<td>-0.6442985</td>
</tr>
<tr>
<td>AASDHPPPT</td>
<td>0.8071383</td>
<td>0.91551061</td>
<td>-0.7459808</td>
<td>-0.9766682</td>
</tr>
<tr>
<td>AADAC</td>
<td>0.8645997</td>
<td>0.83776414</td>
<td>-1.0765764</td>
<td>-0.6257964</td>
</tr>
<tr>
<td>AABT</td>
<td>1.1556595</td>
<td>0.47758562</td>
<td>-1.0431379</td>
<td>-0.5901072</td>
</tr>
<tr>
<td>AAA4</td>
<td>0.4724367</td>
<td>1.04024800</td>
<td>-1.2810233</td>
<td>-0.2316614</td>
</tr>
<tr>
<td>AADAACL2</td>
<td>1.3011498</td>
<td>0.02106791</td>
<td>-1.1261009</td>
<td>-0.1961170</td>
</tr>
<tr>
<td>AAMP</td>
<td>1.1004181</td>
<td>0.59274116</td>
<td>-0.7999765</td>
<td>-0.8931828</td>
</tr>
<tr>
<td>AANAT</td>
<td>1.1659982</td>
<td>0.49625758</td>
<td>-0.7516622</td>
<td>-0.9105936</td>
</tr>
<tr>
<td>AASDH</td>
<td>-0.4675273</td>
<td>1.32468488</td>
<td>-1.0027167</td>
<td>0.1455591</td>
</tr>
<tr>
<td>AAK1</td>
<td>1.0835176</td>
<td>0.59586852</td>
<td>-0.6653927</td>
<td>-1.0139935</td>
</tr>
<tr>
<td>AADAACL4</td>
<td>0.5993446</td>
<td>1.07065287</td>
<td>-1.0586534</td>
<td>-0.6113440</td>
</tr>
<tr>
<td>AADAT</td>
<td>0.6561996</td>
<td>1.04082847</td>
<td>-0.6966487</td>
<td>-1.0083793</td>
</tr>
<tr>
<td>ACAD11</td>
<td>-1.1108356</td>
<td>-0.58004971</td>
<td>0.8409657</td>
<td>0.8499196</td>
</tr>
<tr>
<td>VRK1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSTM4</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VPS9D1-AS1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSRL1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>ACAP1</td>
<td>0.5392450</td>
<td>-1.13710621</td>
<td>0.9127669</td>
<td>0.7635843</td>
</tr>
<tr>
<td>VSTM5</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSIG10L2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSIG10</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSTM1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VRK2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSTM2L</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSIR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSTM2B</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSIG4</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VRK3</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSIG1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSIG10L</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSTM2A</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VPS9D1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSX1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

---

DExMA package
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AASDH</td>
<td>1.28122621</td>
<td>0.6882994</td>
<td>0.3313130</td>
<td>-1.5532911</td>
</tr>
<tr>
<td>AAK1</td>
<td>0.03726305</td>
<td>0.7816216</td>
<td>0.9843186</td>
<td>-1.2179787</td>
</tr>
<tr>
<td>AAACL4</td>
<td>1.09122261</td>
<td>0.8224126</td>
<td>0.2253374</td>
<td>-1.4373346</td>
</tr>
<tr>
<td>AADAT</td>
<td>0.55307845</td>
<td>0.5457234</td>
<td>1.4479847</td>
<td>-1.0401099</td>
</tr>
<tr>
<td>ACAD11</td>
<td>-1.16714015</td>
<td>-0.5365638</td>
<td>-0.8363435</td>
<td>0.3055333</td>
</tr>
<tr>
<td>VRK1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSTM4</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VPS9D1-AS1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSNL1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>ACAP1</td>
<td>-1.17329107</td>
<td>-0.3798016</td>
<td>-1.0565983</td>
<td>0.8999814</td>
</tr>
<tr>
<td>VSTM5</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSIG10L2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSIG10</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSTM1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VRK2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSTM2L</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSI1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSI1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSI10L</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSTM2A</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VPS9D1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>VSTM1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>ACAP1</td>
<td>-1.42596732</td>
<td>-0.4403868</td>
<td>1.3999718</td>
<td>-0.837794528</td>
</tr>
<tr>
<td>AADRCL3</td>
<td>-0.93417370</td>
<td>-1.0076946</td>
<td>1.1791022</td>
<td>0.472019038</td>
</tr>
<tr>
<td>AARS1</td>
<td>-0.17388068</td>
<td>-0.9217414</td>
<td>0.7049468</td>
<td>0.900420939</td>
</tr>
<tr>
<td>AAAS</td>
<td>-1.10680832</td>
<td>-0.2961094</td>
<td>0.8055885</td>
<td>0.825278291</td>
</tr>
<tr>
<td>AASDH</td>
<td>-1.10680832</td>
<td>-0.2961094</td>
<td>0.8055885</td>
<td>0.825278291</td>
</tr>
<tr>
<td>AACS</td>
<td>-0.67777542</td>
<td>-0.6805857</td>
<td>0.3448361</td>
<td>1.245414822</td>
</tr>
<tr>
<td>AAADHPPT</td>
<td>-0.51632041</td>
<td>-0.5997614</td>
<td>1.2322820</td>
<td>0.353638199</td>
</tr>
<tr>
<td>AACD</td>
<td>-0.89449227</td>
<td>-1.3275379</td>
<td>0.3168962</td>
<td>0.927391951</td>
</tr>
<tr>
<td>AARSD1</td>
<td>-0.77018577</td>
<td>-0.5593615</td>
<td>0.6718124</td>
<td>1.01801637</td>
</tr>
<tr>
<td>AABT</td>
<td>-1.42799917</td>
<td>0.9179220</td>
<td>0.8656974</td>
<td>0.849364871</td>
</tr>
<tr>
<td>AAA4</td>
<td>-0.96655055</td>
<td>-0.5743799</td>
<td>0.1665761</td>
<td>1.238929034</td>
</tr>
<tr>
<td>AADACL2</td>
<td>-1.42596732</td>
<td>-0.4403868</td>
<td>1.3999718</td>
<td>-0.837794528</td>
</tr>
<tr>
<td>AAPM</td>
<td>-0.02038311</td>
<td>-1.2776440</td>
<td>0.7285864</td>
<td>0.893817981</td>
</tr>
<tr>
<td>AAANAT</td>
<td>-0.57603553</td>
<td>-1.2334222</td>
<td>1.3990270</td>
<td>-0.46859455</td>
</tr>
<tr>
<td>AASDH</td>
<td>-0.17029375</td>
<td>-0.5772538</td>
<td>0.6129201</td>
<td>0.910894848</td>
</tr>
<tr>
<td>AAK1</td>
<td>-1.23178212</td>
<td>0.6465576</td>
<td>0.9894857</td>
<td>0.371213632</td>
</tr>
<tr>
<td>AADACL4</td>
<td>0.26967066</td>
<td>-0.9713088</td>
<td>1.1607444</td>
<td>-0.834476495</td>
</tr>
<tr>
<td>AADAT</td>
<td>-0.54991192</td>
<td>-0.9567647</td>
<td>0.5509839</td>
<td>0.777842738</td>
</tr>
<tr>
<td>ACAD11</td>
<td>0.94328094</td>
<td>1.2912332</td>
<td>-1.0427073</td>
<td>-0.659717646</td>
</tr>
<tr>
<td>VRK1</td>
<td>NA</td>
<td>NA</td>
<td>-0.3991328</td>
<td>-1.132808807</td>
</tr>
<tr>
<td>VSTM4</td>
<td>NA</td>
<td>NA</td>
<td>1.1480953</td>
<td>-0.843804656</td>
</tr>
<tr>
<td>VPS9D1-AS1</td>
<td>NA</td>
<td>NA</td>
<td>-1.2947008</td>
<td>-0.222935977</td>
</tr>
<tr>
<td>VSNL1</td>
<td>NA</td>
<td>NA</td>
<td>-0.7343571</td>
<td>-0.877211096</td>
</tr>
<tr>
<td>ACAP1</td>
<td>0.64129122</td>
<td>1.0684184</td>
<td>-1.0551468</td>
<td>-0.606268006</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>AARS2</td>
<td>-1.193097649</td>
<td>-0.39494560</td>
<td>1.2047919</td>
<td>1.07007824</td>
</tr>
<tr>
<td>AACDL3</td>
<td>-0.716438370</td>
<td>-0.93468287</td>
<td>0.6082930</td>
<td>0.47422102</td>
</tr>
<tr>
<td>AARS1</td>
<td>-0.353412568</td>
<td>-1.25195516</td>
<td>0.5683881</td>
<td>0.99032956</td>
</tr>
<tr>
<td>AAAS</td>
<td>-0.403674596</td>
<td>-1.22779222</td>
<td>0.5868315</td>
<td>0.60634862</td>
</tr>
<tr>
<td>AACS</td>
<td>-0.993913293</td>
<td>-0.67460343</td>
<td>0.7438120</td>
<td>0.70095867</td>
</tr>
<tr>
<td>AACSP1</td>
<td>-0.690542772</td>
<td>-0.89722647</td>
<td>1.3253285</td>
<td>0.34768914</td>
</tr>
<tr>
<td>AAGAB</td>
<td>-0.614044900</td>
<td>-0.97620692</td>
<td>1.7246362</td>
<td>0.01073741</td>
</tr>
<tr>
<td>AASDH</td>
<td>-1.015277940</td>
<td>-0.57064316</td>
<td>0.9699261</td>
<td>0.67584429</td>
</tr>
<tr>
<td>AADAC</td>
<td>0.173278628</td>
<td>-1.41756675</td>
<td>0.8166646</td>
<td>0.83547946</td>
</tr>
<tr>
<td>AARS1</td>
<td>-0.605348342</td>
<td>-1.07726574</td>
<td>0.3208675</td>
<td>1.54083274</td>
</tr>
<tr>
<td>AABT</td>
<td>-1.028452101</td>
<td>-0.68610202</td>
<td>1.0582122</td>
<td>1.01834320</td>
</tr>
<tr>
<td>AAA4</td>
<td>-1.176875375</td>
<td>-0.22862980</td>
<td>0.6756384</td>
<td>0.98541331</td>
</tr>
<tr>
<td>AADACL2</td>
<td>-0.916553115</td>
<td>-0.44562416</td>
<td>0.6482716</td>
<td>1.08819763</td>
</tr>
<tr>
<td>AAMP</td>
<td>-1.232006994</td>
<td>-0.39039735</td>
<td>1.0541916</td>
<td>0.87742615</td>
</tr>
<tr>
<td>AANAT</td>
<td>-0.004861139</td>
<td>-0.92475211</td>
<td>0.4283357</td>
<td>1.18895237</td>
</tr>
<tr>
<td>AASDH</td>
<td>-1.325546900</td>
<td>-0.19746268</td>
<td>0.3936066</td>
<td>0.46554875</td>
</tr>
<tr>
<td>AAK1</td>
<td>-1.372219271</td>
<td>0.01151991</td>
<td>0.8914460</td>
<td>0.71351602</td>
</tr>
<tr>
<td>AADACL4</td>
<td>-1.276311403</td>
<td>0.15004352</td>
<td>0.6127288</td>
<td>1.15154039</td>
</tr>
<tr>
<td>AADAT</td>
<td>-1.441744415</td>
<td>0.11291774</td>
<td>1.4996941</td>
<td>0.68539206</td>
</tr>
<tr>
<td>ACAD1D</td>
<td>0.732036111</td>
<td>0.97039437</td>
<td>-0.1564368</td>
<td>-0.49161788</td>
</tr>
<tr>
<td>VRK1</td>
<td>1.264157191</td>
<td>0.32778445</td>
<td>-0.9183288</td>
<td>-0.22605111</td>
</tr>
<tr>
<td>VSTM4</td>
<td>1.125981608</td>
<td>0.46591839</td>
<td>-1.8484634</td>
<td>-0.17431453</td>
</tr>
<tr>
<td>VPS9D1-AS1</td>
<td>1.006243659</td>
<td>0.51139316</td>
<td>-0.3766663</td>
<td>-1.71358002</td>
</tr>
<tr>
<td>VSNL1</td>
<td>0.362701612</td>
<td>1.24886662</td>
<td>-0.7663241</td>
<td>-0.88793936</td>
</tr>
<tr>
<td>ACAP1</td>
<td>0.567121587</td>
<td>1.09429317</td>
<td>-0.3591057</td>
<td>-0.91685493</td>
</tr>
<tr>
<td>VSTM5</td>
<td>1.285962931</td>
<td>0.29841958</td>
<td>-0.5761074</td>
<td>-0.25540357</td>
</tr>
<tr>
<td>VSIG10L2</td>
<td>0.682990288</td>
<td>0.72791243</td>
<td>-0.5669168</td>
<td>-0.27272316</td>
</tr>
<tr>
<td>VSIG10</td>
<td>0.498889938</td>
<td>1.14164815</td>
<td>-1.0549086</td>
<td>-0.36552545</td>
</tr>
<tr>
<td>VSTM1</td>
<td>1.010411751</td>
<td>0.65221327</td>
<td>-0.3253520</td>
<td>-1.22683282</td>
</tr>
<tr>
<td>VRK2</td>
<td>0.290643408</td>
<td>1.22584984</td>
<td>-0.2475666</td>
<td>-0.75203655</td>
</tr>
<tr>
<td>VSTM2L</td>
<td>0.612048448</td>
<td>0.98298788</td>
<td>-1.3242282</td>
<td>-0.30928837</td>
</tr>
<tr>
<td>VSIR</td>
<td>0.647524480</td>
<td>1.03758593</td>
<td>-0.7367610</td>
<td>-0.66790259</td>
</tr>
<tr>
<td>VSTM2B</td>
<td>1.011994595</td>
<td>0.55429844</td>
<td>-0.2874520</td>
<td>-1.45039745</td>
</tr>
<tr>
<td>VSIG4</td>
<td>0.862633571</td>
<td>0.82108827</td>
<td>-1.1848497</td>
<td>-0.76428048</td>
</tr>
<tr>
<td>VRK3</td>
<td>0.776871260</td>
<td>0.94918271</td>
<td>-1.3310580</td>
<td>-0.56329900</td>
</tr>
<tr>
<td>Gene</td>
<td>Sample17.Study4</td>
<td>Sample18.Study4</td>
<td>Sample19.Study4</td>
<td>Sample20.Study4</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>VSIG1</td>
<td>1.103462931</td>
<td>0.58574140</td>
<td>-0.6755072</td>
<td>-1.18598708</td>
</tr>
<tr>
<td>VSIG10L</td>
<td>0.474938718</td>
<td>1.12257892</td>
<td>-0.4928980</td>
<td>-1.07710114</td>
</tr>
<tr>
<td>VSTM2A</td>
<td>0.757903621</td>
<td>0.95510300</td>
<td>-0.8669201</td>
<td>-0.72768484</td>
</tr>
<tr>
<td>VPS9D1</td>
<td>0.657559123</td>
<td>1.06494410</td>
<td>-1.0661534</td>
<td>-0.04791842</td>
</tr>
<tr>
<td>VSX1</td>
<td>1.232555953</td>
<td>0.31877803</td>
<td>-0.7673579</td>
<td>-1.42403271</td>
</tr>
<tr>
<td>AARS2</td>
<td>0.33548921</td>
<td>-0.978168251</td>
<td>-0.8478691</td>
<td>-0.78432197</td>
</tr>
<tr>
<td>AADACL3</td>
<td>1.38821537</td>
<td>-1.29633970</td>
<td>-0.8194346</td>
<td>-0.35496081</td>
</tr>
<tr>
<td>AARS1</td>
<td>0.74282169</td>
<td>-0.047906853</td>
<td>-0.5778748</td>
<td>-1.67575772</td>
</tr>
<tr>
<td>AAAS</td>
<td>1.21907318</td>
<td>-0.053486390</td>
<td>-1.2717939</td>
<td>-1.08697301</td>
</tr>
<tr>
<td>AACS</td>
<td>1.08356763</td>
<td>-1.35036645</td>
<td>-0.9520291</td>
<td>-0.22577249</td>
</tr>
<tr>
<td>AAGAB</td>
<td>0.71612384</td>
<td>-0.436915497</td>
<td>-1.4827170</td>
<td>-0.46950899</td>
</tr>
<tr>
<td>AACS</td>
<td>0.12387329</td>
<td>-0.90015143</td>
<td>0.1287143</td>
<td>-1.08780978</td>
</tr>
<tr>
<td>AASDHPI</td>
<td>0.9579171</td>
<td>-1.09178114</td>
<td>-1.2211976</td>
<td>-0.27512806</td>
</tr>
<tr>
<td>AADACL1</td>
<td>0.52121887</td>
<td>-1.087312823</td>
<td>-0.9866706</td>
<td>-0.30893567</td>
</tr>
<tr>
<td>AABT</td>
<td>0.48724030</td>
<td>-0.977729998</td>
<td>-0.3503127</td>
<td>-1.23575392</td>
</tr>
<tr>
<td>AA4</td>
<td>1.26080969</td>
<td>-0.790252699</td>
<td>-0.7292882</td>
<td>-1.16759179</td>
</tr>
<tr>
<td>AADACL2</td>
<td>0.79068870</td>
<td>-1.024505524</td>
<td>-1.2952874</td>
<td>-0.20736496</td>
</tr>
<tr>
<td>AAMP</td>
<td>0.61802876</td>
<td>-0.782743132</td>
<td>-0.3527996</td>
<td>-1.41403767</td>
</tr>
<tr>
<td>AANAT</td>
<td>0.92838867</td>
<td>-0.766724415</td>
<td>-1.3206648</td>
<td>-0.45888756</td>
</tr>
<tr>
<td>AASDH</td>
<td>1.58695527</td>
<td>-0.670733809</td>
<td>-1.0926837</td>
<td>-0.68269299</td>
</tr>
<tr>
<td>AAK1</td>
<td>0.96135491</td>
<td>-0.241840999</td>
<td>-1.2461008</td>
<td>-1.07710606</td>
</tr>
<tr>
<td>AADACL4</td>
<td>0.84468751</td>
<td>-1.244625253</td>
<td>-0.4527872</td>
<td>-0.91154424</td>
</tr>
<tr>
<td>AADAT</td>
<td>-0.24171822</td>
<td>-0.035805706</td>
<td>-0.5054236</td>
<td>-1.40123857</td>
</tr>
<tr>
<td>ACAD11</td>
<td>-1.4649621</td>
<td>0.124966311</td>
<td>1.5398884</td>
<td>0.44816300</td>
</tr>
<tr>
<td>VRK1</td>
<td>-1.27953965</td>
<td>0.944625717</td>
<td>0.2227733</td>
<td>1.24248951</td>
</tr>
<tr>
<td>VSTM4</td>
<td>0.01327709</td>
<td>0.837086850</td>
<td>0.8731125</td>
<td>0.29930145</td>
</tr>
<tr>
<td>VPS9D1-AS1</td>
<td>-0.17621280</td>
<td>1.012295278</td>
<td>0.8351452</td>
<td>0.41301863</td>
</tr>
<tr>
<td>VSNL1</td>
<td>-0.96501288</td>
<td>0.388287090</td>
<td>0.9517297</td>
<td>1.27929593</td>
</tr>
<tr>
<td>ACP1</td>
<td>-0.06453924</td>
<td>1.948688579</td>
<td>-0.4582433</td>
<td>-0.15894545</td>
</tr>
<tr>
<td>VSTM5</td>
<td>-1.28372603</td>
<td>1.437254262</td>
<td>0.9131807</td>
<td>-0.23519795</td>
</tr>
<tr>
<td>VSIG10L2</td>
<td>-1.51565527</td>
<td>0.608446824</td>
<td>0.4200656</td>
<td>1.32678363</td>
</tr>
<tr>
<td>VSIG10</td>
<td>-0.90635013</td>
<td>1.163428345</td>
<td>1.2531779</td>
<td>-0.0892205</td>
</tr>
<tr>
<td>VSTM1</td>
<td>-0.47355971</td>
<td>0.701930653</td>
<td>1.6049881</td>
<td>-0.28567422</td>
</tr>
<tr>
<td>VRK2</td>
<td>-1.48979178</td>
<td>0.651594331</td>
<td>1.0616402</td>
<td>0.77687049</td>
</tr>
<tr>
<td>VSTM2L</td>
<td>-0.66769822</td>
<td>0.004892528</td>
<td>0.9632726</td>
<td>1.33394965</td>
</tr>
<tr>
<td>VSIR</td>
<td>-1.25441145</td>
<td>0.664738911</td>
<td>1.1119140</td>
<td>0.87642214</td>
</tr>
<tr>
<td>VSTM2B</td>
<td>-0.68584179</td>
<td>0.331013623</td>
<td>1.1948646</td>
<td>0.89781307</td>
</tr>
<tr>
<td>VSI4</td>
<td>-0.59129234</td>
<td>0.830268073</td>
<td>0.3497454</td>
<td>1.36040909</td>
</tr>
<tr>
<td>VRK3</td>
<td>-0.64249280</td>
<td>0.994628857</td>
<td>0.3713171</td>
<td>1.17091190</td>
</tr>
<tr>
<td>VSIG1</td>
<td>-0.79024789</td>
<td>0.686231980</td>
<td>1.2173862</td>
<td>0.74812392</td>
</tr>
<tr>
<td>VSIG10L</td>
<td>-1.02829017</td>
<td>0.429639636</td>
<td>1.1256359</td>
<td>1.04301371</td>
</tr>
<tr>
<td>VSTM2A</td>
<td>-1.09845817</td>
<td>0.971509565</td>
<td>1.0688633</td>
<td>0.65269384</td>
</tr>
<tr>
<td>VPS9D1</td>
<td>-1.34019192</td>
<td>0.658610706</td>
<td>0.8004777</td>
<td>0.95175262</td>
</tr>
<tr>
<td>VSX1</td>
<td>-0.33785216</td>
<td>1.144663968</td>
<td>0.6788051</td>
<td>0.70577373</td>
</tr>
</tbody>
</table>
4 Additional information

**DExMA** provides some functions which may be useful for the user, although they are not essential to perform meta-analysis.

### 4.1 GEO microarray data download

In addition to using own user data, **DExMA** package allows to make use of public microarray data from the NCBI GEO public database [3]. For doing that, we can make use of `downloadGEOData()` function. This function uses internally **GEOquey** package in order to download some files at the same time. This function has an input a character vector (`GEOobject`) with the GEO ID of the different datasets that we want to download and a character (directory) that indicates the directory where GSE Series Matrix files [9] are going to be stored.

```r
> GEOobjects<- c("GSE4588", "GSE10325")
> dataGEO<- downloadGEOData(GEOobjects)
```
Once the download process is completed, we get a list of ExpressionSets. This list can be used as input of `createObjectMA()` function, although it is advisable to homogenize gene annotation, in case genes IDs are not Entrez, Official Gene Symbol or Ensembl.

4.2 Using RNA-Seq data

DExMA internally uses `limma` package in order to assess differential expression. Therefore, RNA-Seq data must be previously normalized by the user in order to be able to include correctly these data in the gene-expression meta-analysis. Since `limma` is used internally, we recommend to apply the steps described in the limma user’s guide for the RNA-Seq data normalization [10], although the users can use the type of normalization they prefer.

4.3 Removing Batch Effects

Before the creation of the `objectMA`, a batch effect correction can be applied in order to reduce the effect of covariates that may be affecting to gene expression [1]. Firstly, with functions `prince()` and `prince.plot()` function of the `swamp` package [11], we can obtain a visualization of the p-values of each principal component associated with the categorical covariates.

```r
> library(swamp)
> pheno <- listPhenodatas$Study2
> pheno <- pheno[, apply(pheno, 2, function(x) length(table(x))) > 1]
> # Character variables must be converted in numeric
> pheno <- data.frame(apply(pheno, 2, factor), stringsAsFactors = TRUE)
> res_prince <- prince(listMatrixEX$Study2, pheno, top=ncol(listMatrixEX$Study2))
> prince.plot(res_prince, note=TRUE, notecex=0.5)
```
The categorical variables that may be causing a batch effect can be corrected by using the `removeBatch()` function. The input of this function is the expression matrix and the phenodata. In addition, we also have to add a formula with the variables for which we want to correct the gene expression and the name of the variable that contains the cases and controls groups. Finally, if there is a covariate inside the formula that we want to give greater importance, we will have the option to indicate in the function (`mainCov()`). Here we show an example in which we have corrected the gene expression of the previous study by two of their covariates:

```r
> listMatrixEX$Study2 <- batchRemove(listMatrixEX$Study2, listPhenodatas$Study2, formula=~gender+race, mainCov = "race", nameGroup="condition")
```

Coefficients not estimable: batch1 batch2 (Intercept) raceC raceH

```r
> head(listMatrixEX$Study2)
Sample5  Sample6  Sample7  Sample8  Sample9  Sample10
100859927 4.367690 6.648252 5.143704 2.820073 2.688687 2.688687
8086 5.937004 5.914434 5.831974 3.234439 3.228559 3.228559
8212 5.294553 4.092231 2.575022 3.368330 2.750491 2.750491
65985 5.685822 4.892231 2.750222 3.368330 2.750491 2.750491
729522 5.733054 5.553966 6.798683 3.275042 2.659061 2.659061
```
4.4 Calculating Effects size

The `calculateES()` function returns the effects size in each of the studies. Moreover, it calculates the variance of each of the effects and the \( \log_2 \) Fold Change. The effects size are calculated by making use of the *Hedges’ g estimator* [1].

```r
> effects <- calculateES(newObjectMA)
> head(effects$ES)

Study1       Study2       Study3       Study4
AAA4 1.448942 2.5008385 1.1221262 6.642233
AAAS 1.506204 2.9643045 2.2593392 2.430575
AABT 2.288894 0.7656633 5.7273196 3.478798
AACS 2.579019 5.3315818 1.8723418 1.207949
AACSP1 5.705553 5.1083967 2.9008711 3.138604
AADAC 7.515582 2.1970319 0.8345501 3.776154
```

```
> head(effects$Var)
```

![Color Key and Histogram](image)

- **Condition**
  - 0.001
  - 3e-04
  - 1
  - 0.02
  - 0.4
  - 0.2

- **Gender**
  - 1
  - 0.06
  - 1
  - 1
  - 0.3
  - 0.9

- **Race**
  - 2e-05
  - 0.1
  - 1
  - 0.7
  - 0.2
  - 0.2

Principal Components (Variation)
### 4.5 Calculating Individual P-values

Similar to the calculation of effects sizes, the individual p-values of each of the studies and the $\log_2$ fold change of each one can also be calculated by applying `pvalueIndAnalysis()` function. P-value are obtained by assessing differential expression with `limma` package.

```r
> pvalues <- pvalueIndAnalysis(newObjectMA)
> head(pvalues$p)
```

<table>
<thead>
<tr>
<th>Study</th>
<th>p</th>
<th>Study</th>
<th>p</th>
<th>Study</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA4</td>
<td>0.084586213</td>
<td>AAA4</td>
<td>0.0124328637</td>
<td>AAA4</td>
<td>0.137545477</td>
</tr>
<tr>
<td>AAAS</td>
<td>0.078241874</td>
<td>AAAS</td>
<td>0.0065691298</td>
<td>AAAS</td>
<td>0.032924671</td>
</tr>
<tr>
<td>AABT</td>
<td>0.031889060</td>
<td>AABT</td>
<td>0.2765998575</td>
<td>AABT</td>
<td>0.003696266</td>
</tr>
<tr>
<td>AACS</td>
<td>0.024335148</td>
<td>AACS</td>
<td>0.005986238</td>
<td>AACS</td>
<td>0.049748561</td>
</tr>
<tr>
<td>AACSP1</td>
<td>0.003718102</td>
<td>AACSP1</td>
<td>0.0007176307</td>
<td>AACSP1</td>
<td>0.018622696</td>
</tr>
<tr>
<td>AADAC</td>
<td>0.001900842</td>
<td>AADAC</td>
<td>0.0197199607</td>
<td>AADAC</td>
<td>0.223678418</td>
</tr>
</tbody>
</table>

```r
> head(pvalues$logFC)
```

<table>
<thead>
<tr>
<th>Study</th>
<th>logFC</th>
<th>Study</th>
<th>logFC</th>
<th>Study</th>
<th>logFC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA4</td>
<td>2.167998</td>
<td>AAA4</td>
<td>2.654066</td>
<td>AAA4</td>
<td>3.118286</td>
</tr>
<tr>
<td>AAAS</td>
<td>2.931649</td>
<td>AAAS</td>
<td>2.663952</td>
<td>AAAS</td>
<td>4.321073</td>
</tr>
<tr>
<td>AABT</td>
<td>3.857649</td>
<td>AABT</td>
<td>1.030831</td>
<td>AABT</td>
<td>3.259410</td>
</tr>
<tr>
<td>AACS</td>
<td>3.376387</td>
<td>AACS</td>
<td>2.881147</td>
<td>AACS</td>
<td>4.252159</td>
</tr>
<tr>
<td>AACSP1</td>
<td>2.157823</td>
<td>AACSP1</td>
<td>3.164180</td>
<td>AACSP1</td>
<td>3.998437</td>
</tr>
<tr>
<td>AADAC</td>
<td>6.771502</td>
<td>AADAC</td>
<td>2.282112</td>
<td>AADAC</td>
<td>1.747278</td>
</tr>
</tbody>
</table>

### 5 Session info

R Under development (unstable) (2023-11-11 r85510)
Platform: x86_64-pc-linux-gnu
Running under: Ubuntu 22.04.3 LTS
DExMA package

Matrix products: default
BLAS: /home/biocbuild/bbs-3.19-bioc/R/lib/libRblas.so
LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.10.0

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

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

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

other attached packages:
[1] swamp_1.5.1 MASS_7.3-60.1 gplots_3.1.3
[4] amap_0.8-19 impute_1.77.0 Biobase_2.63.0
[7] BiocGenerics_0.49.1 DExMA_1.11.8 DExMADATA_1.11.0

loaded via a namespace (and not attached):
[1] tidyselect_1.2.0 farver_2.1.1 dplyr_1.1.4
[4] blob_1.2.4 Biostrings_2.71.1 bitops_1.0-7
[7] fastmap_1.1.1 RCurl_1.98-1.13 GEOquery_2.71.0
[10] XML_3.99-0.16 digest_0.11.1 lifecycle_1.0.4
[13] Biostrings_1.70.0 survival_3.5-7 statmod_1.4.1
[16] KEGGREST_1.43.0 RSQLite_2.0.2 magrittr_2.0.3
[19] genefilter_1.85.0 compiler_4.0.4 rlang_1.0.6
[22] tools_4.4.0 utf8_1.1.2  igraph_1.7.1
[25] yaml_2.2.7 data.table_1.14.8 knitr_1.45
[28] bit_4.0.5 plyr_1.8.9 xml2_1.3.5
[31] RColorBrewer_1.1-3 KernSmooth_2.23-22 BiocParallel_1.37.0
[34] purrr_1.0.2 grid_4.4.0 stats4_4.4.0
[37] fansi_1.0.5 caTools_1.18.2 xtable_1.8-4
[40] colorspace_2.1-0 edgeR_4.1.2 scales_1.3.0
[43] gtools_3.9.5 cli_3.6.1 rmarkdown_2.25
[46] crayon_1.5.2 generics_0.1.3 tzdb_0.4.0
[49] httr_1.4.7 DBI_1.1.3 cachem_1.0.8
[52] zlibbioc_1.49.0 splines_4.4.0 parallel_4.4.0
[55] AnnotationDbi_1.65.2 BiocManager_1.30.22 XVector_0.43.0
[58] matrixStats_1.1.0 vctrs_0.6.4 Matrix_1.2-8
[61] sva_3.50.0 hms_1.1.3 IRanges_2.27.0
[64] S4Vectors_0.41.2 bit64_4.0.5 locfit_1.5-9.8
[67] limma_3.59.1 tidyr_1.3.0 snpStats_1.53.0
[70] annotate_1.81.0 glue_1.6.2 codetools_0.2-19
[73] gtable_0.3.4 GenomeInfoDb_1.39.1 munsell_0.5.0
[76] tibble_3.2.1 pillar_1.9.0 htmltools_0.5.7
[79] GenomeInfoDbData_1.2.11 R6_2.5.1 evaluate_0.23
References


