

# geod24243

María Lairón y Guillermo Ayala

2024-04-02

## Table of contents

Data . . . . .	1
Packages . . . . .	1
Download data from ArrayExpress . . . . .	1
Robust multichip average . . . . .	2
Making an ExpressionSet . . . . .	2

## Data

The data set can be found at [GEO gse24243](#).

## Packages

```
pacman::p_load(AnnotationDbi, ArrayExpress, mouse4302.db, pd.mouse430.2)
```

## Download data from [ArrayExpress](#)

Using package **ArrayExpress**.

```
geod24243raw = ArrayExpress("E-GEOD-24243")
save(geod24243raw, file=paste0(dirTamiData, "geod24243raw.rda"))
```

```
load(paste0(dirTamiData, "geod24243raw.rda"))
```

Let us see the class.

```
class(geod24243raw)
```

```
[1] "ExpressionFeatureSet"  
attr(,"package")  
[1] "oligoClasses"
```

## Robust multichip average

```
geod24243 = rma(geod24243raw)
```

## Making an ExpressionSet

1. Experimental data.

```
datosexperimento = new('MIAME',name='Flowers MT, Scheisser K, Dawson JA,  
                          Kendziorski C, Ntambi JM',  
                          lab='James Ntambi',  
                          contact='mtflowers@wisc.edu',  
title = 'Metabolic changes in skin caused by Scd1 deficiency: a focus on  
retinol metabolism.',  
abstract = 'To help elucidate the metabolic changes in the skin that  
contribute to the obesity resistance and skin pathology in mice lacking Scd1, we performed  
url = 'https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-24243/'  
)  
experimentData(geod24243) = datosexperimento
```

2. Define a new experimental factor.

```
type = factor(c(0,0,0,1,1,1),levels = 0:1,labels=c("SKO","LOX"))  
pd = data.frame(type)  
rownames(pd) = colnames(geod24243)  
pData(geod24243) = pd
```

3. Annotation

Let us simplify the problem and we are going to consider just the correspondences with ENTREZID.

```
a = AnnotationDbi::select(mouse4302.db, keys=featureNames(geod24243),
                           column=c("ENTREZID"),
                           keytype="PROBEID")
```

#### 4. Multiple correspondences

- How many features have after normalization?

```
nrow(geod24243)
```

We have multiple correspondences.

- How many and which are the probes corresponding with more than one gene.

```
b = which(table(a[,1]) > 1)
length(b)
head(b)
```

- For the probes `names(b)` we have more than one gene.

```
sel = is.element(a[,1], names(b))
table(table(a[sel,1]))
```

- A possibility is to use the first correspondence in the original `data.frame`.

```
c1 = match(unique(a[,1]), a[,1])
a1 = a[c1,]
```

- Let us see if each ENTREZID identifier has a unique PROBEID identifier.

```
table(table(a1[,2]) > 1)
```

- It is not true.
- Let us choose one probe per ENTREZID.

```
c2 = match(unique(a1[,2]), a1[,2])
a2 = a1[c2,]
dim(a2)
```

- Is there a 1-1 correspondence?

```
length(unique(a2[,1])) == length(unique(a2[,2]))
```

- Now we want to establish the correspondence between PROBEID with ENTREZID and ENSEMBL.

```
a = AnnotationDbi::select(mouse4302.db, keys=featureNames(geod24243),
                           column=c("ENTREZID", "ENSEMBL"),
                           keytype="PROBEID")
View(a)
b = match(featureNames(geod24243), a[, "PROBEID"])
fData(geod24243) = a[b,]
```

Let us see the dimension of a.

```
dim(a)

c1 = match(unique(a[,1]), a[,1])
a1 = a[c1,]
c2 = match(unique(a1[,2]), a1[,2])
a2 = a1[c2,]
```

The ENSEMBL identifiers are not unique.

```
table(table(a1[,3]) > 1)
```

- It is important to choose the primary identifier for later use.

4. Saving data.

```
save(geod24243, file = "geod24243.rda")
```