

# GSE25171

Lucía Martínez Priego      Guillermo Ayala

26 de marzo de 2025

## Resumen

It is a study about the response to the phosphate deficiency in Arabidopsis.

## 1 Data

We define the variable to control the execution of the code chunks.  
The directory where the data sets have to be saved.

## 2 Packages

The needed packages are loaded.

```
pacman::p_load(Biobase,GSEABase,GEOquery,annotate,affy,genefilter,GOstats,
               Category,ath1121501.db)
```

## 3 Downloading the data

```
getGEOSuppFiles("GSE25171")
```

Normalizing the data set.

```
setwd(dirTamiData)
system("tar xvf GSE25171/GSE25171_RAW.tar")
gse25171raw = ReadAffy()
gse25171rma = rma(gse25171raw)
system("rm -fr GSE25171")
system("rm GSM*.gz")
```

Arabidopsis grew hydroponically in the presence (control group) as well as in the absence of phosphate (treatment group). The samples were also collected at different times: 0 and 1 hour (short times) and 6 and 24 hours (average times). Three independent replicas were made in each case. The study aims to observe the changes in the pattern of expression between these two times. We add the metadata or phenotypic variables.

```

time = rep(rep(c(0,1,6,24),each=2),3) ## hours
##time = factor(time,levels=c(0,1,6,24),labels=c("0hr","1hr","6hr","24hr"))
time2 = rep(rep(c(0,1),each = 4),3)
time2 = factor(time2,levels=c(0,1),labels=c("Short","Medium"))
Pi = rep(c(1,0),12)
Pi = factor(Pi,levels=c(0,1),labels=c("Control","Treatment"))
replication = rep(1:3,8)
metadatos = data.frame(time,time2,Pi,replication)
rownames(metadatos) = colnames(gse25171)
pData(gse25171) = metadatos

```

The annotation is

```

annotation(gse25171)

```

We are going to add different gene identifiers.

```

keytypes(ath1121501.db)
a = select(ath1121501.db,keytype="PROBEID",keys= featureNames(gse25171),
          columns = c("ENTREZID","GO","TAIR"))

```

We establish a unique correspondence between PROBEID and ENTREZID

→ . First, we remove probes without correspondence with ENTREZID.

```

a = a[-which(is.na(a$ENTREZID)),]

```

```

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

```

Now, we have to select the rows in the ExpressionSet.

```

rows = match(a2[,1],featureNames(gse25171rma))
gse25171 = gse25171rma[rows,]
all(a2$PROBEID==featureNames(gse25171))
fData(gse25171) = a2

```

Saving the ExpressionSet.

```

save(gse25171,file=paste0(dirTamiData,"gse25171.rda"))

```

We construct a gene set collection for arabidopsis thaliana.

```

pacman::p_load("ath1121501.db")
frame = toTable(org.At.tairGO)
goFrameData = data.frame(frame$go_id, frame$Evidence, frame$gene_id)
goFrame = GOFrame(goFrameData, organism = "Arabidopsis")
goAllFrame = GOAllFrame(goFrame)
gscAt = GeneSetCollection(goAllFrame, setType = GOCollection())
gscAt = geneIds(gscAt)
save(gscAt,file = paste0(dirTamiData,"gscAt.rda"))

```